ABSTRACT

Title of Dissertation: KNOWLEDGE AND PROCESSING OF MORPHOSYNTACTIC VARIATION IN AFRICAN AMERICAN LANGUAGE AND MAINSTREAM AMERICAN ENGLISH

Zachary Kevin Maher
Doctor of Philosophy, 2023

Dissertation Directed by: Professor Jan Edwards
Associate Professor Jared Novick
Department of Hearing and Speech Sciences

As people from different social groups come into contact, they must accommodate differences in morphosyntax (e.g., *He seem nice* vs. *He seem**s** nice*) in order to successfully represent and comprehend their interlocutor's speech. Listeners usually have high comprehension across such differences, but little is known about the mechanisms behind morphosyntactic accommodation. In this dissertation, I asked what listeners know about variation in morphosyntax and how they deploy this knowledge in real-time language processing. As a test case, I focused on regularized subject-verb agreement (e.g., *He seem nice*, *They was happy*)—which is common in African American Language (AAL), but not in Mainstream American English (MAE)—and compared how listeners adjust their linguistic expectations depending on what language varieties both they and their interlocutors speak.

In Experiment 1, I showed that participants who primarily speak MAE 1) recognize that some speakers use regularized subject-verb agreement, 2) evaluate that regularized subject-verb agreement is associated with AAL, and 3) predict that the subject-verb agreement rules of AAL

allow for some patterns (*They was happy*) but not others (**He were happy*). This was accomplished using a novel sentence rating task, where participants heard audio examples of a given language variety, then rated written sentences for how likely a speaker of that variety would be to say them. In Experiment 2, I showed that a similar pool of participants did not merely recognize regularized subject-verb agreement; their knowledge of variation lead them to predict that AAL speakers use regularized forms in an acoustically ambiguous context. Participants heard sentences like *He sit(s) still*, where it is unclear whether the verb includes a verbal *-s* due to a segmentation ambiguity. They were more likely to transcribe a regularized form (*He sit still*) when it was spoken by an AAL-speaking voice than when it was spoken by an MAE-speaking voice. Together, these results indicate that listeners have rich mental models of their interlocutors that extend beyond a general awareness of linguistic difference.

In Experiment 3, I compared bidialectal speakers of AAL and MAE and monodialectal speakers of MAE. On the rating task from Experiment 1, bidialectal participants showed a greater degree of differentiation between sentences that are grammatical in AAL and sentences that are ungrammatical in AAL, compared to monodialectal participants. However, both groups of participants indicated that ungrammatical sentences are broadly more likely in AAL than MAE, contrary to usage patterns in the world. On the transcription task from Experiment 2, bidialectal participants were overall more likely to transcribe regularized subject-verb agreement, but they differentiated between AAL- and MAE-speaking voices to the same degree as monodialectal participants. Both groups were more likely to use MAE subject-verb agreement (*He sits still*) than regularized subject-verb agreement (*He sit still*). These results suggest that bidialectal listeners broadly expect regularized subject-verb agreement to a greater degree than do monodialectal listeners, rather than making stronger predictions about a given speaker. Moreover, while bidialectal listeners have a more granular sense of AAL's grammatical rules, all listeners still favor

MAE, likely reflecting MAE's dominant status.

In Experiment 4, I asked how listeners use their knowledge of variation in subject-verb agreement to guide real-time interpretation of sentences, again comparing bidialectal and monodialectal participants. Participants heard sentences like *The duck(s) swim in the pond*, where they must rely on the agreement morphology of the verb to determine whether the subject of the sentence is singular or plural, since a segmentation ambiguity makes it unclear whether the noun ends in *-s*. In MAE, only a plural interpretation is available, while in AAL, a singular interpretation is also available. Participants' eye-movements were tracked as they looked at and selected images on a screen. Participants were more likely to look at and select a singular image if the sentence was presented in an AAL-speaking voice, compared to an MAE-speaking voice, and bidialectal participants were more likely to look at and select a singular image, compared to monodialectal participants. As with the transcription task in Experiment 3, this suggests that bidialectal participants are broadly more likely to consider the possibility that a speaker uses regularized SVA, compared to monodialectal participants, but their linguistic expectations are not more strongly differentiated based on the grammar of their interlocutor.

These results make it clear that listeners have mental models of morphosyntactic variation, which can be characterized along a variety of dimensions, including the syntax, semantics, and indexicality (social meaning) of a given variable. This can serve as a foundation for future inquiry into the details of these models and the real-time switching and control dynamics as listeners adjust to different varieties in their environment.

KNOWLEDGE AND PROCESSING OF MORPHOSYNTACTIC VARIATION IN AFRICAN AMERICAN LANGUAGE AND MAINSTREAM AMERICAN ENGLISH

by

Zachary Kevin Maher

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Jan Edwards, Co-chair
Associate Professor Jared Novick, Co-chair
Assistant Professor Shenika Hankerson
Professor Jeffery Lidz, Dean's Representative
Professor Colin Phillips
Assistant Research Professor Charlotte Vaughn

Acknowledgments

I am deeply grateful for the immense support I have received in pursuing this project and completing my path to a Ph.D. I would never have reached this point without many gifted teachers and mentors, supportive family and friends, and people working behind the scenes who made all this work possible. I cannot begin to adequately thank everyone, and there are many important people who I have not listed here, but I truly value everyone in the community of people in my life who made this possible.

I can't imagine more supportive advisors than Jan Edwards and Jared Novick. Thanks to both of you for showing me what it looks like to pursue important questions, while also maintaining a constant openness to new ideas and a constant care for everyone in your life. Thank you for encouraging me in this work even as it took me to new questions beyond what your labs had been doing and for lifting me up through all the setbacks of building a new research program during a pandemic. Thank you to the members of my committee—Shenika Hankerson, Jeff Lidz, Colin Phillips, and Charlotte Vaughn—for the support and thoughtful suggestions that have made this dissertation stronger, and for believing in this project in its most nascent stages as I found the most productive ways to pose my research questions. Working with all of you has been a reminder of why I love interdisciplinary language science!

Thank you to everyone else who made this project practically possible. Thank you to the National Science Foundation for making this work made financially possible (NSF-DDRI #2215102, NSF-NRT #1449815) and to the administrative staff who guided me through all of the documentation for using these resources. Thank you to the speakers for putting up with

hours of recording sessions to get the stimuli in these experiments just right, to Rochelle Newman and her lab for use of the recording booth, and to participants for the time completing these experiments. Thank you to Yi Ting Huang for allowing me to use your eye-tracker. Thanks to Maddie Lenhart for all of your work on the eye-tracking study. It was a pleasure working with you, and it was exciting to see you take on the project and make it your own. I am also thankful to audiences at Psychonomics, the American Dialect Society, and Human Sentence Processing conferences for feedback and encouragement on parts of this work.

Thank you to Patrick Plummer, Shenika Hankerson, Arynn Byrd, Charity Ayebae, Temptaeous Mckoy, for help recruiting bidialectal participants. An extra thank you to Patrick Plummer for allowing me to run participants in your lab space. Not only was this a massive help for data collection; you made it fun to be around for what would have been grueling days of running participants!

I have spent my graduate training in a rich interdisciplinary language science community, and this has profoundly shaped how I do research and engage as a scholar. Thank you to everyone who has kept the Maryland Language Science Center going, including (but not limited to) Colin Phillips, Shevaun Lewis, Caitlin Eaves, and Tess Wood. Thank you to the many smaller groups of language scientists who have taken me in, including the psycholinguistics check-in group and GAMMs club.

Thank you to my labmates for support and feedback throughout the years and the development of this project, both in the Learning to Talk Lab (Christina Blomquist, Allie Johnson, Tatiana Thonesavanh, Michelle Erskine, Sidra Sheikh, Arynn Byrd, Kathleen Oppenheimer, Meg Cychosz, Carolyn Mazzei) and Language and Cognition Lab (Tal Ness, Zoe Ovans, Lauren Salig, Kelly Marshall), as well as our collaborators at CU Boulder. A particular thank you to Christina and Allie for your friendship as we all got adjusted to a new place and to Arynn and Michelle for

formative conversations about studying language variation.

Finally, thanks to everyone outside of my academic community who has helped me reach this point. Thanks to my parents for your consistent love and support, for feeding my love of learning from an early age, for years of helping me to revise my writing (which made this large writing task possible!), and for providing a physical space to organize the ideas that led to this dissertation while the world was in lockdown. To my sister Megan, for both moral support and editing support on this document. And to Sara: thank you for persevering through these six years of long distance; your love and companionship have provided constant grounding and joy, even through the stress of lockdowns and dissertation writing. Can't wait to see what comes next!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAL	African American Language
AAVE	African American Vernacular English
ERP	Event-related potential
EEG	Electroencephalography
CORAAL	Corpus of Regional African American Language
MAE	Mainstream American English
SVA	subject-verb agreement

Chapter 1: Introduction

Imagine a typical introductory Linguistics course. Throughout the course, the instructor will reiterate the idea that the goal of the linguist is to be *descriptive*, characterizing people's mental grammars as they are, rather than *prescriptive*, making recommendations for the "right" way to speak. There will typically be a dedicated unit on sociolinguistics. One goal of this unit will be to show that all language varieties (or "dialects") are governed by grammatical rules; common societal narratives postulating that some people acquire "bad grammar" actually reflect prejudice and broader oppressive structures in society (Lippi-Green, 1997; Baker-Bell, 2020). In teaching this, the instructor will often draw upon common examples of stigmatized forms, anticipating that the students will recognize the forms, but not know the grammatical rules. This practice suggests an assumption that the students—often speakers of Mainstream American English (MAE)—bring some knowledge of variation with them to the classroom, but that this knowledge is incomplete. However, this assumption has received surprisingly little theoretical and experimental attention. In this dissertation, I will ask what listeners know about morphosyntactic variation, with an eye toward laying the foundation for a larger theoretical framework.

Let us begin with some common examples of morphosyntactic variation. Speakers of MAE and speakers of African American Language (AAL)[1] can generally understand each other, but

[1]I am using these terms because they currently seem to be the dominant terms in the literature. As many have noted, much diversity is contained within both labels. "Mainstream" or "Standard" American English tends to be defined by the absence of stigmatized variants (Wolfram and Schilling, 2015). See, for example, Lippi-Green's (1997) definition of "standard" language: "an abstract, idealized homogeneous language, which is imposed and maintained by dominant institutions and which has as its model the written language, but which is drawn primarily from the spoken language of the upper middle class" (p. 64). The term *African American Language* is deliberately broad,

they speak differently in reliable ways. While an MAE speaker might produce the forms in (1), an AAL speaker might produce the forms in (2):

(1) a. He **is** working.

b. He work**s**.

(2) a. He ∅ working.

b. He work∅.

c. He **be** working.

From the perspective of the MAE speaker, the sentences in (2) are ungrammatical but largely interpretable, though (2c) might be erroneously taken to be closer in meaning to (1a) than (1b). Conversely, someone who produces the sentences in (2) would likely find (1b) interpretable and might even produce (1a) in some circumstances. In most circumstances, these differences do not lead to changes in propositional meaning; there is no context in which (2b) would have different truth conditions[2] from (1b). In other cases, there can be subtle differences in truth conditions, potentially creating uncertainty or false certainty on the part of the listener. For example, since *be* is used for habitual aspect in AAL, the correct answer to (3) is (a), though an MAE speaker might incorrectly guess (b) (Jackson, 1998):

(3) Who be eating cookies? (Jackson, 1998)

a. Cookie monster (currently not eating)

b. Elmo (currently eating a cookie)

It makes intuitive sense that listening to a less familiar language variety could be difficult,

rather than centering the language of working-class Black male speakers (King, 2020). Most of the phenomena that I discuss would also fall under the narrower label of *African American Vernacular English* (AAVE).

[2]Throughout this dissertation, I occasionally use the terms "propositional meaning" or "truth conditions," since various other aspects of meaning, including the social meaning, might differ between variants, even when what is narrowly asserted about the world is the same.

but the mechanistic underpinnings of this difficulty remain unclear and understudied. As Terry et al. (2010, 2015) argue, one possibility is that accommodating unfamiliar morphemes imposes demands on working memory. A broader version of this account might be that listeners must edit any sentence to conform with their own grammar, and this process depletes processing resources. Alternatively, listeners might reduce their reliance on particular morphosyntactic cues in processing when their interlocutor speaks an unfamiliar variety. This would reduce any cost associated with editing "anomalies," but also reduce the amount of redundant cues available in processing. These hypotheses are not mutually exclusive; listeners might use different strategies depending on their knowledge of their interlocutor's grammar. However, relatively little is known about listeners' knowledge of grammatical variation, particularly of variants that they do not produce. In this dissertation, I implement novel methods to (1) more precisely characterize listener knowledge of the grammatical and social conditions of subject-verb agreement (SVA) variation in U.S. English and (2) test how listeners deploy their knowledge of SVA to guide real-time interpretation.

In this chapter, I will begin with a discussion of the broader societal significance of the processing of dialect variation (Section 1.1). I will then outline a framework by Labov (1973) to understand different types of knowledge that an individual might have about the grammar of a different dialect (Section 1.2). I will then consider existing work on sociolinguistic processing (Section 1.3), relevant work in psycholinguistics on the processing of anomalous utterances (Section 1.4.1), and work on dialect-contingent processing strategies (Section 1.4.2). I conclude with a general outline of this dissertation (Section 1.5) and notes on my positionality (Section 1.6).

1.1 Applied consequences

While subtle differences in the processing of SVA might seem like a trivial task, there has been mounting evidence suggesting that differences in morphosyntax can pose a challenge in some

circumstances. This has been most thoroughly explored for children who speak AAL. A large body of work has found correlations between the degree to which a child's language differs from MAE and lower scores on a variety of literacy measures (e.g., Charity et al., 2004; Gatlin and Wanzek, 2015). While most of this work appeals to general dialect differences, Terry and colleagues have argued for a specific effect of verbal *-s*: the presence of verbal *-s* in mathematical word problems imposes a cognitive load, decreasing performance on these tasks. Terry et al. (2010) coded test items from the Woodcock-Johnson-R Applied problems subtest, which consists of mathematical word problems, for six morphosyntactic features: past tense *-ed*, participle *-en*, past tense copula, auxiliary *have*, 3rd person singular *-s*, and counterfactual conditional *if*...*-ed*. They modeled correct responses to a given question as the dependent variable, and the only feature with a significant effect was verbal *-s*. The correlation became stronger when children's AAL production rate was included in the model.

Terry et al. (2015) report that this effect also showed the greatest variability in its effect on student scores: "In the highly affected group, roughly 15% of participants, the average student would answer 9 percent more questions correctly if the effect of this feature were removed" (pp. 645-646). While this work is limited by a correlational design, Terry et al. (2022) conducted follow-up experiments manipulating the presence or absence of *-s*, again finding a significant result for overt verbal *-s*; children performed 10% worse with an overt marking relative to zero marking. The special status of verbal *-s* is consistent with other work on AAL. Early descriptive work, such as that of Labov (1969), argued that verbal *-s* is categorically absent in AAL (4), unlike a feature like null copula, which is variable, with overt copulas also being acceptable in AAL (5) (Labov, 1969):

(4) a. He work∅ from home on Fridays. (AAL: ✓, MAE: *)

b. He work**s** from home on Fridays. (AAL: *, MAE: ✓)

(5) a. She ∅ working from home right now. (AAL: ✓, MAE: *)

b. She **is** working from home right now. (AAL ✓, MAE: ✓)

This finding has largely been corroborated (Green, 2011, 2019; Cleveland and Oetting, 2013), though there have been exceptions (e.g., Baugh, 1979, 1990; Ball, 1995). It is also consistent with findings that AAL-speaking children do not interpret verbal *-s* in on-line tasks (de Villiers and Johnson, 2007). Terry et al. (2022) built on these findings using electroencephalography (EEG) to study AAL-speaking 2nd graders' on-line processing of verbal *-s*. They found early anterior negativity for verbal *-s* relative to null marking, which could be interpreted as an indication that the *-s* in sentences like (4b) are ungrammatical to these participants. When the subject of the sentence was plural, the authors also observed a late central-posterior positivity, which they interpret as a P600, potentially indicating error repair. This is consistent with a suggestion by Labov and Baker (2015) that AAL-speaking children attempt to incorporate the unfamiliar morpheme as a plural *-s*.

Though most work has focused on minoritized speakers' comprehension of MAE, MAE speakers can also have poor performance when listening to AAL. In the classroom environment, we might think of the role of a predominantly White teaching workforce with limited training in AAL and negative attitudes toward its speakers (e.g., Cross et al., 2001; Godley et al., 2015). Beyond the classroom, court reporters have accuracy rates that fall far below accepted thresholds when transcribing sentences of AAL exhibiting a range of morphosyntactic phenomena (Jones et al., 2019). While Jones et al. did not focus their analysis on individual phenomena, other studies have found that adult MAE speakers have low rates of correct interpretation for stressed *BIN* (6, Beyer et al., 2015) and habitual *be* (Wolfram, 1982).

(6) Chad *BIN* using that cell phone (Beyer et al., 2015)

'Chad has been using that cell phone for a long time.'

Taken together, these findings show that linguistic challenges compound broader social justice issues facing Black speakers of AAL: speakers of AAL must contend with the difficulties associated with comprehending MAE, as well as the potential to be misunderstood by MAE speakers wielding institutional power. There are many possible interventions that might make sense to combat this problem, including curricula that affirm minoritized varieties (Wheeler, 2016), training for teachers (Mallinson et al., 2011) and court reporters, and broader initiatives aimed at fighting systemic racism and classism (Lewis, 2018).

Though this dissertation does not directly test interventions, it is an attempt to understand the processing of variation at the level of individual cognition. This could help us to explain why given approaches to training might be successful. Instead of focusing attention on teaching minoritized speakers to use MAE (e.g., Wheeler, 2016), this dissertation primarily places the burden of cross-variety adaption on people who primarily speak MAE, as this is the more dominant group. Thus, it is more oriented toward interventions on MAE-dominant teachers and court reporters than toward interventions on AAL-speaking students.

1.2 Types of knowledge of morphosyntactic variation

The above work on applied consequences of dialect differences appeals to a general notion of familiarity, but it is clear that individuals who produce different forms come into contact with each other and often have social judgments about some linguistic variants. Despite this, there has been relatively little work on listeners' knowledge of morphosyntactic variation. Labov (1973) provides a useful framework for characterizing different possible types of knowledge possible for a sociolinguistic morphosyntactic phenomenon. For a given speaker A' of Dialect A, he outlines different types of knowledge of rule B_1 of Dialect B:

> There are six questions we can pose about A's grasp of a rule of Dialect B which will help us decide what kind of a rule he himself [*sic*] is using. A' has just heard a speaker

> of B use a form B_1 that is not produced by A''s grammar. (1) Does A' *recognize* B_1 as grammatical for some native speakers of English? In a word, does he know that B_1 exists? Or, faced with B_1 for the first time, can he recognize it as a possibility open to speakers of English? (2) Can he *evaluate* its social significance——that is, see it as colloquial, formal, slang, or stigmatized so that he would know in what social context to use it? (3) Can he *interpret* B_1—not just in the normal favorable contexts, where it is supported by other forms, but in neutral and unfavorable contexts as well? (4) Can A' *label* the meaning of B_1 in zero contexts, faced with the isolated form itself? ... (5) With or without this understanding, can A' *predict* the use of B_1 in an extended range of environments——both syntactic and semantic? Finally, (6) can he *use* B himself productively? (6) does not automatically follow from (5), since the use of language in social interaction requires a much higher degree of skill and a kind of overlearning that is not needed to predict the use of others in a reflective mode. (pp. 45-46)

For example, speakers of AAL might use habitual *be* (*He be working*), which is used to indicate habitual actions. Following this framework, an MAE speaker who does not produce this form might exhibit varying levels of knowledge: (1) *recognizing* that some people use this pattern; (2) *evaluating* that it is associated with AAL; (3) *interpreting* the habitual aspect (e.g., completing the cookie monster task in example 3); (4) *labeling* this meaning; (5) *predicting* the morphosyntactic constraints on the phenomenon; and (6) *using* the phenomenon. In the final case of *using* habitual *be*, we might say that the speaker in question is bidialectal, speaking both MAE and AAL. Depending on participant group, Labov found mixed levels of knowledge from participants who would not produce a range of phenomena including negative concord, positive *anymore*, and aspectual phenomena in AAL. Generally high levels of knowledge of negative concord have been replicated in a recent series of studies (Blanchette and Lukyanenko, 2019a,b).

Though *recognition* is a logically necessary precondition for all other parts of Labov's framework, the remaining types of knowledge—indexical, semantic, and morphosyntactic—could be independent. For example, a given speaker might be able to correctly predict the morphosyntactic constraints on a phenomenon without evaluating its indexical value for a particular social group. This has been found for a phonological variable in Austen's (2020) dissertation: even listeners who

know the linguistic constraints of the TRAP/BATH split in British English will assign the social category of "posh" to *incorrect* contexts for the BATH vowel (e.g., GAS, a lexical exception).

Additionally, there are cases where individuals correctly predict morphosyntactic constraints for a phenomenon they have probably not encountered. Wolfram (1982) tested knowledge of the linguistic conditions of habitual *be* (see 2c above), as well as *a*-prefixing (7) among middle class native English speakers in the Washington, D.C. area. Participants were presented with pairs of sentences (e.g., 7) and selected the sentence where either *be* or *a-* could be included. They were highly accurate with the *a*-prefixing items, but not the habitual *be* items.

(7) a. He went a-sailin'.

b. *He likes a-sailin'. (Wolfram, 1982)

This is surprising, since these participants were more likely to have exposure to AAL than Appalachian English. Wolfram cautiously reasons that participants were generalizing from other linguistic constraints, which themselves might have influenced the linguistic conditions for *a*-prefixing. He does not report participants' beliefs about which social groups use each phenomenon, but it is plausible that under Labov's framework, participants *evaluated* that habitual *be* is associated with AAL, but could not accurately *interpret* its meaning or *predict* its morphosyntactic constraints; on the other hand, they could *predict* the morphosyntactic constraints on *a*-prefixing without necessarily knowing that this phenomenon is associated with Appalachian English. In Wolfram's (1982) study, there was no difference in truth conditions between forms with and without *a*-prefixing, but when these factors can be separated, a speaker might also correctly predict the morphosyntactic constraints of a phenomenon without understanding its meaning. For example, one might correctly identify the grammaticality constraint for habitual *be* illustrated in (8) without understanding the truth conditions of the habitual aspect. (To my knowledge, this has not been tested.)

(8) a. I never be looking for that.

b. *I be never looking for that. (Green, 2002)

In addition to the separation between the *type of information* being represented (morphosyntactic, semantic, indexical), there is a second dimension to knowledge of variation: *granularity* of the knowledge, which would include factors such as the ability to interpret an utterance with or without supporting context and sensitivity to grammatical conditions. This idea of granularity has been previously noted in Preston's (1996) account of folk sociolinguistic knowledge, which focuses on knowledge of phonological variation. He notes that folk descriptions vary in both their *accuracy* and their level of *detail*. Level of detail might range from "global," where the presence of a different accent is noted without any linguistic particulars, to "specific," where a specific feature is noted, such as "g-dropping." There is also evidence of listener sensitivity to grammatical conditioning, indicating a high degree of granularity. Vaughn and Kendall (2018) found that comprehenders were sensitive to the probabilistic grammatical constraints on (ING) variation, even in a task where they explicitly rated how surprising an *-in* realization would be. Most speakers of American English have both (ING) variants in their repertoire, so it remains to be seen how sensitive such rating measures will be for forms that they do not produce.

Variation in subject-verb agreement (SVA) presents an interesting test case here. Unlike (ING), SVA rules are often categorical between dialects (see discussion in Section 1.1), and SVA phenomena are more purely morphosyntactic, but there is still a plausible range of granularity of knowledge. If speaker A' of Dialect A is listening to SVA differences in Dialect B, they might (1) know that Dialect B has broadly different grammatical rules, (2) know that Dialect B has different SVA rules, or (3) know the specific SVA paradigm of Dialect B. In the case of an MAE speaker listening to AAL, the listener might expect all of the below sentences, (9b) and (9c), or only (9c), respectively.

(9) a. *Him was happy.

b. *He were happy.

c. They was happy.

While this dissertation will mostly examine the granularity of morphosyntactic knowledge, granularity differences could also apply to semantic and indexical knowledge. Much work in sociolinguistics focuses on the specifics of the social judgments that listeners make when they hear a particular variant, and research on indexical fields (Eckert, 2008) shows that speakers also express rich social identities through combinations of features.

A final dimension under consideration in Labov's framework is people's *conscious awareness* of a given form, which he terms *labeling*. Preston (1996) refers to a similar idea under the term "availability," where a given phenomenon can range from "unavailable," where a non-linguist is unable to describe a form, even when prompted, to "available," where a phenomenon is subject to "usual folk linguistic discussion." This dimension has also received more recent attention from Squires (2016), who adapts a framework from Schmidt (1990) for second language learning making a distinction between "perceiving" and "noticing." She reports evidence from a self-paced reading study where reading time effects were observed for variation in SVA patterns regardless of whether participants reported that they noticed the manipulation following the task; however, there were complex interactions with participants' awareness of the manipulation. Furthermore, Campbell-Kibler (2016) relates questions of listener awareness to the process of making social inferences from linguistic variables, noting parallels to dual-systems models in social psychology split between "fast" and "slow" processing. For this dissertation, I am primarily interested in underlying linguistic representations and the process which Squires (2016) would describe as "perceiving," rather than the metalinguistic processes that Labov describes as "labeling." Despite this, metalinguistic measures still bear on the questions at hand, since metalinguistic knowledge

presupposes linguistic knowledge. This aligns with Labov's point that "the labelling function might seem beside the point, since the normal use of language does not require it; but it will appear that in the heartland of Dialect B, speakers will be able to choose the right label when outsiders fail" (p. 46).

1.3 Knowledge of the social meaning of sociolinguistic variables

Most of the work on sociolinguistic perception examines participants' social inferences based on the linguistic context, which approximately corresponds to indexical knowledge, or the *evaluation* part of Labov's framework. For example, in the *sociolinguistic monitor* account, listeners track frequencies of features with a social evaluation in order to make judgments about their interlocutor, and the evidence for this account comes from alveolar vs. velar realizations of the (ING) variable (Labov et al., 2011; Campbell-Kibler, 2011). For example, Labov et al. (2011) found that as the proportion of *-in* tokens increased in text, listeners were less likely to rate the speaker as being well-suited for a job as a newscaster.

Labov et al. deliberately focused on sociophonetic variation due to the *interface principle*, which asserts that "members of the speech community evaluate the surface form of language but not more abstract structural features," meaning that "surface" phonetics are more appropriate for sociolinguistic study than "deep" morphosyntax (Labov et al., 1993). However, Levon and Buchstaller (2015) found that participants in the United Kingdom made reliable social evaluations based on the presence of SVA patterns that differ from Standard British English (10),[3] as well as a sociophonetic variable (11), indicating that morphosyntactic variation is also subject to social evaluation.

(10) They really like**s** ice-cream. (Levon and Buchstaller, 2015)

[3]Note that British variation in SVA is different from SVA variation in the United States.

(11) I **f**ink [think] we should go home now (Levon and Buchstaller, 2015)

Squires (2014) found comparable effects for SVA variation in U.S. English. She studied listeners' associations of *don't* with singular subjects (a common variant in U.S. English) and *doesn't* with plural subjects (uncommon in U.S. English). She found that participants associated both forms with images of someone with a lower socioeconomic status, but with increased reaction times in the case of the uncommon plural+*doesn't*. The methodology differs in that participants had to make a binary judgment between two images, but it offers confirmation that speakers in the U.S. do socially evaluate SVA. It further raises the question of granularity, suggesting that listeners broadly associate differences in SVA with a lower socioeconomic status, but they also have more specific expectations for the attested form.

Work by Bender (2005) also finds a relationship between granularity of morphosyntactic knowledge and social evaluation. She found that listeners in the United States associated the null copula in AAL with a variety of social characteristics, such as level of education, and this finding varied by listener background. For the participants who were familiar with AAL, there was an interaction between the linguistic environment and their social evaluation. In AAL, null copulas appear more frequently preceding predicates with V+*ing* (12a) than an NP (12b) (Labov, 1969):

(12) a. She ∅ teachin me piano at Music World.

b. She ∅ my piano teacher at Music World. (Bender, 2005)

The latter, more marked, form triggered a stronger social evaluation (lower ratings for how "educated" the speaker sounded) for participants familiar with AAL, but not for other participants. Similarly, Vaughn (2022) found that listeners gave higher accentedness ratings to a speaker who used the *-in* variant in a rare grammatical environment (e.g., *everythin*) relative to a common environment (e.g., *runnin*), though this was only true for non-spliced stimuli. This suggests that

there is an interaction between detailed knowledge of a given variety (*predicting* syntactic details in Labov's framework) and social evaluation, where less typical forms are viewed as less "standard," but only if the listener can identify the forms as particularly marked.

1.4 Processing of variation

Since one goal of this dissertation is to draw connections between sociolinguistic knowledge and processing, it is worth considering how Labov's framework might apply in real-time processing before reviewing the existing literature on sociolinguistic processing. Thus far, I have discussed knowledge of variation along two dimensions: the type of knowledge and the granularity of the knowledge. When considering questions of sentence processing, there is an additional dimension to consider: the *time course* in which different types of knowledge of variation become available to a listener. It is possible that a classic performance-competence difference would emerge with regard to differences in dialects, where off-line tests indicate greater knowledge of D_B than on-line processing. For example, A' might ignore SVA when processing D_B but still accurately report the specific patterns of D_B, or A' might make incremental interpretations according to D_A's grammar, but with greater readiness to revise when confronted with new information. Depending on a listener's experience, it might be more or less optimal to make specific predictions, depending on the likelihood that those predictions will need to be revised (Valdés Kroff and Dussias, 2023). In this section, I review evidence on real-time processing of morphosyntactic variation.

1.4.1 Accommodation of morphosyntactic anomalies in comprehension

While sociolinguistic variables carry a social signal for evaluation, they are a potential source of noise in processing. One potential way to characterize the task of listening to the morphosyntax of a relatively unfamiliar variety is that of anomaly detection. As noted above, a sentence like *He*

work (2b) might be anomalous to a speaker of MAE due to the absence of verbal *-s*; conversely, *He works* (1b) might be anomalous to a speaker of AAL, even though the meaning of the sentence is easily recovered.

There is extensive work in the field of psycholinguistics on the processing of morphosyntactically anomalous utterances, which often highlights apparent differences between linguistic performance and linguistic competence. Experimental materials in psycholinguistic studies are often intended to be anomalous to participants, meaning that the materials differ from participants' typical experience of language. For example, work on agreement attraction involves sentences like (13), which participants rate as acceptable and process with minimal difficulty, despite agreement patterns that are not licensed by their grammar:

(13) The key∅ to the cabinet**s** *are rusty. (Bock and Miller, 1991)

The reasons for this are a source of ongoing debate, with varying accounts about the role of memory (Christianson, 2016; Parker, 2019). Interestingly, in some cases, participants actually arrive at an incorrect final interpretation where *key* is plural (Patson and Husband, 2016). This finding has not been interpreted in relation to sociolinguistic variation in SVA, but it could indicate that when faced with differences in SVA, listeners will less reliably encode/retrieve whether the subject of a sentence is singular or plural.

There are two popular approaches to explaining why listeners arrive at interpretations that do not fully align with the syntax of the sentence: good enough processing and noisy channel modeling. According to the good enough processing approach (Ferreira et al., 2002), listeners often rely on heuristics in processing and sometimes fail to revise when these heuristics lead them astray. While it is plausible that listeners are more reliant on heuristics when processing a less familiar dialect, I do not think that this approach would make specific predictions about processing across dialects. According to the noisy channel approach (Gibson et al., 2013), fallibility in the

system is not a "bug" stemming from usually-efficient shortcuts in morphosyntactic processing; it is a "feature," where listeners' goal is to recover intended meanings with the knowledge that the speaker's intended meaning could be corrupted. Thus, if a listener perceives an implausible sentence like (14a), they will likely arrive at a non-literal interpretation, inferring that the speaker intended a minimally-different but plausible alternative (14b).

(14) a. The mother gave the candle the daughter.

b. The mother gave the candle **to** the daughter.

Several factors might affect the likelihood that listeners make a so-called noisy channel inference. For present purposes, the most important factor is the likelihood that a sentence contains an error; if errors are likely, a listener should be more likely to infer a non-literal intended meaning. This proves true when a high proportion of written experimental items contain anomalies (Gibson et al., 2013) and when items are recorded in a non-native accent (Gibson et al., 2017). The latter finding makes sense under the assumption that listeners expect more errors from non-native speakers than from native speakers (though it is unclear whether listeners truly interpret dialect differences as errors, even if they would describe them as such for ideological reasons). Additionally, listeners are sensitive to the specific types of noise that are likely to appear, rather than relying on an overall error rate. Ryskin et al. (2018) manipulated whether participants were exposed to deletion errors, addition errors, a mixture, or exchange errors, and found that for both deletion errors and exchange errors, participants adapted to the particular error type.

Adjustment to types of errors is also reflected in the event-related potential (ERP) literature, where the P600 effect has been associated with morphosyntactic anomalies. This effect is characterized by a late posterior positivity, around 500-900 ms following the presentation of an anomalous stimulus. For example, in sentences like (15), Molinaro et al. (2008) observed a P600

following the anaphor *themselves*, which is anomalous (in MAE) following the number features of both the subject and verb.

(15) The famous dancer was nervously preparing herself/*themselves to face the crowd. (Molinaro et al., 2008)

The precise characterization of the P600 remains a source of ongoing research and debate, but a common view is that it indexes a repair process, particularly when a small edit would rectify an anomaly (Kim and Sikos, 2011; Ryskin et al., 2021). This is relevant to the study of morphosyntactic variation because the P600 effect is sensitive to the expected anomalousness of the experimental context. Hahne and Friederici (1999) manipulated the proportion of sentences in the experiment and found a P600 effect for sentences with phrase structure violations when there was a low proportion of sentences with such violations, but no such effect when there was a high proportion of anomalous sentences. Hanulíková et al. (2012) found a similar result when manipulating the accent of a speaker; with native Dutch-speaking participants, the researchers observed a P600 effect for grammatical gender for native-accented Dutch but not for Turkish-accented Dutch. Grey and Van Hell (2017) also found a reduced P600 effect when native U.S. English speakers heard pronoun errors in Chinese-accented speech, though there were possible interactions with experience with the accent; listeners who successfully identified the accent showed an N400-like response to grammatical errors in Chinese-accented speech, but listeners who did not identify the accent showed no ERP effect.

Weissler and Brennan (2020) extended this work to the processing of dialect variation present in the United States. Participants heard sentences in MAE or AAL, where critical items were grammatical in both dialects (16a), grammatical only in AAL (16b), or ungrammatical in both dialects (16c). A P600 effect was observed for sentences spoken by an MAE-speaking voice, but only in the ungrammatical condition. No P600 effect was found for any sentence type

spoken by an AAL-speaking voice. A potential issue here is that all sentences were recorded by the same bidialectal speaker, so participants might not have consistently differentiated the dialects. In a follow-up study using a White MAE speaker, a P600 effect was observed for both MAE-ungrammatical conditions (Weissler, 2021).

(16) a. She **is** blushing

b. She ∅ blushing

c. *She**'ll** blushing (Weissler and Brennan, 2020)

In addition to reduced P600 effects, reading time measures generally increase following morphosyntactic violations. Squires (2014) found that participants were slower to read sentences with non-standard *don't* with a singular subject (*He don't*) than the MAE-compatible alternative (*He doesn't*). However, this effect was reduced by attributing the text to pop songs, where non-standard forms are more expected (Squires, 2019). Additionally, Kaschak and Glenberg (2004) found that participants showed an initial reading time cost when processing the unfamiliar *needs washed* construction (e.g., *My car needs washed*), but this effect was reduced over the course of the experiment as participants gained familiarity with the construction.

There is mixed evidence about whether listeners are sensitive to the particular types of "anomalies" that are likely in the speaker's dialect/accent. In the above studies, Squires (2014, 2019) found even slower reading times for the essentially unattested (in her terms, "uncommon") option of plural subjects with *doesn't* (*They doesn't*), indicating some latent familiarity with variation in the United States. Caffarra and Martin (2019) and Zhou et al. (2019) found a reduced P600 effect for common errors, but not for uncommon errors (in English-accented Spanish and Chinese-accented English, respectively). However, Weissler and Brennan (2020) found that listeners failed to show a P600 when ungrammatical sentences like (16c) were uttered by an AAL speaker, even though such sentences are ill-formed in AAL.

Overall, evidence from P600 and reading time studies suggests that listeners do not constantly repair morphosyntactic "anomalies" in less familiar linguistic contexts. Rather, in some circumstances, listeners reduce their reliance on morphosyntactic cues in processing. In Labov's framework of knowledge, they are *evaluating* that a particular language variety is associated with grammatical anomalies, but not accurately *predicting* the typical limits of the anomalies. In other circumstances, they have more specific predictions. Such results are not contradictory; they indicate that different listener populations will adjust their processing strategies based on their knowledge of a particular variety and phenomenon.

1.4.2 Influences of sociolinguistic context and listener knowledge on perception

A handful of studies provide preliminary evidence that listener knowledge and sociolinguistic context can interact to shape perception and comprehension of the linguistic signal. Engel and Hanulíková (2020) studied variation between genitive and dative case in German, where there is an ongoing change in which dative case is replacing genitive in some contexts. In these contexts, dative case is more likely to be used than genitive case in casual speech (compared to careful speech) and by younger speakers (compared to older speakers). The authors used a sociolinguistic repetition task (Buson et al., 2018), where participants heard a sentence in careful or casual speech, with case morphology masked by white noise. Participants were more likely to repeat the sentence using the dative case morphology when it was presented in casual speech than in careful speech, and this effect was slightly stronger for younger participants. This indicates that listeners are more likely to perceive a variant that is aligned with the sociolinguistic context, at least when both variants are in their linguistic repertoire.

Lundquist and Vangsnes (2018) studied speakers of Norwegian with varying dialect profiles, exploiting a difference between dialects with a three-gender system (masculine, feminine, neuter)

and a more dominant two-gender system (masculine and feminine are merged). Using the visual world paradigm, participants listened to a word in a sentence where it was preceded by a gender-marked determiner, and the researchers asked whether participants used gender information in predicting the upcoming target. Distinct profiles emerged: participants who spoke only the two-gender dialect only made use of the gender markers found in their native dialect; participants who spoke the three-gender dialect made reliable gender-based predictions, and "stable" users of this dialect adjusted their predictions based on the speaker, while "unstable" users—whose productions were variable between a two- and three-gender system—did not. This suggests that differences in the processing of morphosyntactic variation depend on linguistic experience.

Turning to U.S. English, Garcia et al. (2022) did not find a P600 effect for sentences with regularized SVA (17) for bidialectal speakers of AAL and MAE, but monodialectal speakers of MAE did show a P600 effect. This indicates that regularized SVA was anomalous to monodialectal participants, but not to bidialectal participants. Zaharchuk et al. (2021) observed different results for multiple modals (18). This phenomenon is present in Southern U.S. English, and the authors compared participants who spoke Southern U.S. English or a more general variety that does not include this phenomenon. All participants—including Southern speakers—showed an early anterior negativity and P600, indicating in-the-moment expectations that align with MAE. However, Southern participants gave higher ratings to these sentences in an offline task, revealing some difference in knowledge of these forms. In both studies, all sentences were recorded by Southern White speakers, so these results do not tell us how listeners adjust their expectations depending on who their interlocutor.

(17) The black cat **lap** the milk. (Garcia et al., 2022)

(18) She thinks she **might should** ask the professor for an extension. (Zaharchuk et al., 2021)

There has been more work examining the effect of phonological variation on lexical processing, including studies where aspects of the speaker's dialect was manipulated. Staum Casasanto (2009) presented audio clips where there was a potential lexical ambiguity, depending on the presence of a phonological rule for consonant cluster reduction (e.g., [mæs] for *mass* or *mast*). Participants were faster to select a completion of the sentence that aligned with the reduced interpretation when the sentence was spoken by a Black voice than by a White voice.[4] King and Sumner (2014) used a similar logic in a lexical priming paradigm with a lexical decision task; participants showed greater priming effects for both consonant cluster reduction and TH-fronting for primes presented in a AAL voice, compared to primes presented in an MAE voice. Notably, there was no difference in the dialect effect for cluster reduction compared to TH-fronting, even though cluster reduction occurs in both dialects, while TH-fronting is limited to AAL; this suggests that dialect effects on language processing do not entirely correspond to linguistic patterns in the world. Sumner and Samuel (2009) also studied dialect-contingent effects of lexical priming, comparing the priming effects on non-rhotic productions associated with NYC English to rhotic productions associated with General American English. For participants from NYC, priming effects were observed for both rhotic ([beɪkɚ] for *baker*) and non-rhotic ([beɪkə]) forms, while for non-NYC participants, only rhotic forms were primed. For both groups, however, there were strong priming effects for the rhotic (non-NYC) variant, suggesting a cognitively distinct role for the socially dominant form, regardless of a given listener's primary dialect.

While these studies examine effects after an entire word has been presented, Dahan et al. (2008) found that adjustment to a speaker's dialect occurs incrementally, changing what cohort is under consideration. When participants were exposed to an accent where the /æ/ vowel raises before voiced consonants (*bag*) but not voiceless consonants (*back*), they used this information

[4]I am following APA style in capitalizing both "Black" and "White" for racial identity categories, but there are substantial differences in the type of identity construct and history behind these labels.

to make earlier looks to the correct target when the vowel was a useful cue. Additionally, the linguistic context affects recognition of sociolinguistic variables. In Vaughn and Kendall (2018), listeners pressed a button to indicate whether they heard *-ing* or *-in*, varying whether the preceding context matched or mismatched the production of the informal variant. Participants were faster to select the *-in* variant when the preceding context was matched than when it was mismatched. Overall, these studies make it clear that listeners do adjust their linguistic expectations based on their model of their interlocutor's dialect, but relatively little is known about how they do this for *morphosyntactic* variables.

1.5 Summary and open questions

Morphosyntactic variation is prevalent, but as we have seen, it is poorly understood from the perspective of individual speakers' cognition. Further work is important for both theoretical and applied reasons. From a theoretical perspective, adaptation to morphosyntactic variability is a key component of language processing, but existing models are not designed to account for variation that does not impact propositional meaning. Sociolinguistic variation provides an avenue to study the interaction between listeners' previous linguistic knowledge and their exposure to variability in an experimental setting. From an applied perspective, there is accumulating evidence that morphosyntactic variation can be a particular source of challenge for speakers of minoritized varieties. To understand the particular sources of challenge, we must first characterize the processes that listeners use to adapt to variability.

Evidence from early work in sociolinguistics and more recent work in psycholinguistics shows that listeners can show knowledge of variants that they do not produce and adjust processing based on the social context (genre, speaker identity) and linguistic context. Results have shown mixed levels of knowledge in the Labov (1973) framework. For example, Wolfram (1982)

found that MAE speakers could *predict* the conditions for *a*-prefixing but not habitual *be*, and Squires (2019) found that participants showed a specific advantage for processing *don't* with singular subjects when attributed to song lyrics (prediction and evaluation), but Weissler and Brennan (2020) found a general reduced expectation of grammaticality in AAL. These results also leave open questions regarding comprehension, since the "anomalies" under study did not impact propositional meaning, and comprehension question responses were generally at ceiling. This means that these studies do not show us how listeners would *interpret* the morphosyntactic differences without redundant cues.

At this point, it is difficult to draw broad generalizations from the existing work, since each study examines one snapshot of knowledge for one or two variables, usually with only one participant group. In order to evaluate the boundaries of listeners' grammars in a more systematic way, it is necessary to test multiple types of knowledge with multiple participant groups. In this dissertation, I will focus on regularized subject-verb agreement (SVA), comparing different types of knowledge across different participant groups: monodialectal speakers of MAE and bidialectal speakers of MAE and AAL. I begin with the hypothesis that monodialectal speakers of MAE will recognize that SVA differences are associated with AAL, and reduce their reliance on SVA morphology when listening to AAL. I predicted that bidialectal speakers would interpret and predict the grammatical conditions of SVA for both varieties, and use SVA morphology in comprehension according to their interlocutor's grammar.

In Chapter 2, I develop two paradigms for testing participants' knowledge of morphosyntactic variation and use these paradigms to show that participants who predominantly use MAE show some knowledge of regularized SVA. In Chapter 3, I use these paradigms to compare listeners' knowledge of SVA grammatical patterns across monodialectal and bidialectal participants. In Chapter 4, I again compare monodialectal and bidialectal participants, testing how they use

their knowledge of SVA variation to guide interpretation in real time. Chapter 5 serves as a general discussion and conclusion, evaluating the effect of linguistic experience on knowledge and processing of morphosyntactic variation, with the goal of building toward future frameworks that expand beyond Labov's (1973) initial approach to the question.

1.6 Positionality

In all research—and particularly questions that relate to matters of social identity—it is important to consider how the researcher's identity and experience shape all phases of inquiry. I am a White, cisgender, male researcher from a middle class family, and I grew up in the northeastern United States. English is my first language, and growing up, I only spoke MAE at home. I speak no other variety of English, and my accent is minimally marked; I have never experienced linguistic discrimination. These aspects of my identity have undoubtedly shaped the way I have framed my research questions and designed my experiments. While I seek to frame knowledge of multiple varieties as a valuable skill and to place the burden of change on dominant speakers like me, I recognize the risk that at times, I have been reductive in my approach to AAL speakers and encourage any criticism on this (and any other factors that I do not even anticipate!). In addition to shaping how I posed research questions and analyzed data in this dissertation, my identity also shaped my interactions with participants and likely affected how they approached each task. I will consider this throughout my analyses, particularly in Chapter 4.

Chapter 2: Measuring knowledge of variation in subject-verb agreement

As I discussed in Chapter 1, there has been relatively little work on listeners' knowledge of morphosyntactic variation, including in the common area of SVA, though work by Squires (2014) suggests that listeners from a university pool show some sensitivity to regularization in *doesn't/don't*, showing a reduced reading time cost when regularized forms are attributed to song lyrics. In this chapter, I develop two approaches to measure listeners' knowledge of variation in SVA, with two goals. First, on a theoretical level, I am asking what MAE speakers know about regularized SVA. Second, on a methodological level, I am testing whether two new paradigms are sensitive to different aspects of listeners' knowledge.

In Experiment 1, I asked whether MAE speakers explicitly *recognize* regularized SVA and *evaluate* that it is associated with AAL. I also probed the *granularity* of their knowledge of AAL's SVA rules. Experiment 1 is a pilot study of a novel sentence rating task that I developed for this dissertation, and I will use versions of this task with all participant groups in subsequent chapters. In Experiment 2, I asked whether listeners implicitly *predict* that regularized SVA occurs *in a neutral context*, using an adaptation of the sociolinguistic repetition task (Buson et al., 2018). Throughout three different versions of the task (A, B, and C), I tested the impact of different design choices on participants' performance on the task, with the goal of determining the best version to use when comparing different groups of participants in Chapter 3.

For these initial studies, I focused on the knowledge of White, mid- to high-SES speakers of U.S. English. Listeners with this type of demographic background probably use MAE most of

the time, meaning that they are unlikely to use a range of common but stigmatized morphosyntactic variants, including regularized SVA. Thus, the present studies asked whether people have knowledge of linguistic forms that they are unlikely to produce. This differs from much of the work on sociolinguistic perception (e.g., Vaughn and Kendall, 2018; Labov et al., 2011), which has focused on variables like (ING), where most members of the prestige speech community use both variants. This is particularly relevant for the sociolinguistic repetition task (Buson et al., 2018) used in Experiment 2, which has only been used with broadly-available variables in French (Buson et al., 2018) and German (Engel and Hanulíková, 2020), but never in English (to my knowledge). To preview the results, for all studies, I found clear evidence that these participants have some knowledge of regularized SVA, both recognizing it when it is presented to them and predicting it in neutral contexts, and they were more likely to attribute regularized SVA to AAL than to MAE, indicating a shared social evaluation.

2.1 Experiment 1: Sentence rating task

In Experiment 1, I developed a novel task to study participants' knowledge of the morphosyntactic patterns of different language varieties. This task was based on a traditional acceptability judgment task, where participants rate the degree to which a sentence sounds acceptable/grammatical on a Likert-type scale. For such tasks, experimental language scientists are always faced with the challenge of making sure participants do not merely respond according to prescriptive grammatical rules, and this becomes an acute challenge when studying stigmatized morphosyntactic variants. With the present task, I pushed this a step further, asking participants to provide an acceptability rating for a speaker of a given dialect, rather than for themselves. I used this task to test the hypothesis that listeners who do not produce AAL morphosyntax will

correctly associate a particular verb morphology[1] phenomenon with AAL.

While this task is novel, an analogous approach was taken by Holliday and Tano (2021). In their task, participants rated whether a tweet was grammatical, and tweets were presented with either a Black or White avatar, with sentences falling into one of three categories: grammatical MAE, grammatical AAL, or ungrammatical "AAL." They found that participants gave higher ratings to Black avatars than white avatars for both grammatical MAE and grammatical AAL, and low ratings for both avatars for ungrammatical AAL items. There was no effect of participant race for grammaticality judgments, though there were some complex interactions between participant race and their ratings for different personality attributes of the "speaker," such as whether they sounded "educated." While these results indicate some broad familiarity with the grammatical rules of AAL, there was no systematic manipulation of grammatical feature, and the authors reason that their results were partially driven by the "socially conscious and liberal views" of their undergraduate participant pool, which drew primarily from prestigious, West coast liberal arts schools. This is interesting from the perspective of the study of language ideology, but it limits our ability to interpret the results in terms of knowledge of morphosyntactic variation; as Campbell-Kibler (2016) notes, "individuals with a commitment to linguistic equity may find ... that they catch themselves drawing stereotype-based conclusion about an interlocutor and attempt to rectify those impressions after the fact" (p. 142). The present task takes a more controlled approach to manipulating the linguistic features in question and to ensuring that participants evaluate their responses relative to the grammar of a dialect of interest.

[1]Materials in this study use both variation in SVA and copula overtness, which is broader than subject-verb agreement. Though Labov (1969) provided a phonological analysis of the copula's distribution, Bender (2000) argues that a syntactic analysis is necessary. Regardless of the analysis, the relevant question here is participants' knowledge of each variable's grammatical conditions.

2.1.1 Methods

2.1.1.1 Participants

Participants ($n = 24$) were recruited through Prolific; 3 were excluded from analysis due to incorrect responses on "catch trials," more than one inaccurate response to audio items within a given guise, and/or an incorrect response to the comprehension question for the task instructions (see below). I used Prolific's demographic pre-screening survey to identify likely "monodialectal" speakers of MAE. The following parameters were selected:

- Location: United States
- Nationality: United States
- Age: 18-50
- Ethnicity: White/Caucasian
- First Language: English
- Socioeconomic Status (SES): 6-10 (out of 10)

The SES rating that Prolific uses is an adapted form of the MacArthur Scale of Subjective Social Status (Adler et al., 2000), which asks participants to rank where they fall on a ladder from "worst off" to "best off" in terms of money, education, and jobs. These broad demographic categories are simplistic and hide many sources of experiences with variation. Though I am limited to self-report, I asked a series of questions at the end of the study to get a sense of participants' experience with linguistic and other diversity. The recruited sample included a range of backgrounds in terms of age (mean=33.4, s.d.=8.4), gender identity (11 male, 10 female), and region of the U.S (see summary by state in Table 2.1). Most had completed a four-year college degree or more (Table 2.2).

Table 2.1: Number of participants who grew up or currently live in each U.S. state for Experiment 1.

State	Childhood	Current
California	3	4
Virginia	3	1
New York	3	0
Pennsylvania	2	3
Florida	2	2
North Carolina	1	2
Arizona	1	1
Connecticut	1	1
Maine	1	1
Michigan	1	1
Nebraska	1	0
New Jersey	1	0
Texas	1	0
Illinois	0	1
Massachusetts	0	1
Minnesota	0	1
Nevada	0	1
South Carolina	0	1

Table 2.2: Education levels, Experiment 1

Level of Education	Number of Participants
Graduate school	7
4-year college degree	9
Associate's/Technical degree, trade school, or some college	2
High school diploma or equivalent	3

Self-reported usage and exposure ratings for AAL and MAE are provided in Figure 2.1. Participants generally reported MAE usage and exposure at the highest point in the scale, and AAL usage in the bottom two levels of the scale. In later chapters, I will compare across different linguistic backgrounds, but for the purpose of a pilot study for this task, we can think of this group as MAE-dominant, with varying amounts of AAL in their repertoire.

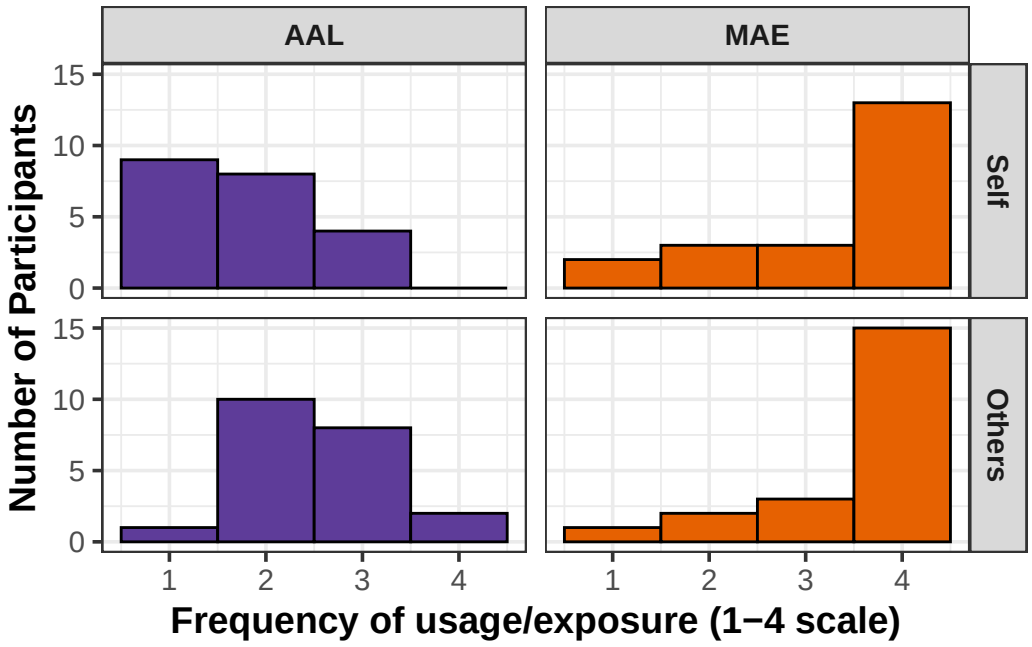

Figure 2.1: Self-reported usage of and exposure to AAL and MAE

2.1.1.2 Procedure

The entire experiment was implemented in Qualtrics. After completing the consent form, participants received general instructions on the task. The exact text was as follows:

> People speak in different ways depending on many different factors, such as their age, profession, and the community where they grew up. This includes different accents, as well as differences in grammar. For example, some people use "double negatives," so they might say "I don't want nothing," while other people might say "I don't want anything." In this study, we want you to listen to examples of different ways of speaking, then rate whether people who speak this way would say certain sentences. Focus on how the sentence is put together, not on what the sentence is about. Don't worry about whether they use the grammar rules that are taught in school. We're interested in how people actually talk in their everyday lives!

With this text still on the screen, they answered the following multiple choice comprehension question:

> How should you rate the sentences in this study?
> A) Based on the "proper" way to speak
> B) Based on how people speak in their everyday lives [Correct answer]
> C) Based on what the sentence is about

The purpose of this question was to force participants to read and understand the instructions, and to allow for the exclusion of participants who did not do so. Additionally, the incorrect answers reflected alternative approaches to the task that I wanted participants to avoid, and this question forced participants to reject those approaches.

After these instructions, participants rated sentences for a given "language variety"; throughout the experiments in this dissertation, I will use the term *guise* to refer to variety/(dia)lect as an experimental manipulation, to distinguish this from dialects as they appear in the world or are spoken by a given participant group. Items were blocked by guise, and block order was counterbalanced across participants. Each block had two components: an introduction phase and a rating phase.

In the **introduction phase**, for each guise, participants first listened to a series of naturalistic audio clips to exemplify the guise. Examples of AAL were drawn from the Corpus of Regional African American Language (CORAAL, Kendall and Farrington, 2021), and examples of MAE were drawn from the Santa Barbara Corpus (Du Bois et al., 2005). Audio examples did not include the phenomena in the experimental items, in order to ensure that participants were drawing on their previous knowledge of each variety, rather than generalizing from the examples provided. AAL examples included negative concord (19), completive *done*, non-rhotic pronunciations, and /f/ where /θ/ would be used in MAE. MAE examples included MAE equivalents of these phenomena (e.g., 20). A complete list of items can be found in Appendix A. Clips were adjusted in Praat to 68 dB,[2] but audio quality varied across clips due to the varied recording environments.

(19) I don't know nobody over there no more.

(20) Now you can't do anything to this guy.

[2]Thank you to Matt Winn for making this script and posting it publicly!

After each audio clip, participants selected one word that appeared in the clip from a three-option multiple choice question. The correct answer was always the site of the feature in question (e.g., for [19], the correct answer was "nobody"). This served the dual purpose of ensuring that participants actually listened to the audio clips and drawing attention to places where there is a clear MAE/AAL distinction. Particpants could also select a fourth option, "I can't understand this clip," to distinguish guessing and random clicking from any intelligibility issues. Audio clips were embedded WAV files, and participants could play them as many times as they wanted to. There were four clips presented per guise in the introduction phase, and then two sets of two items at break points throughout the rating phase to serve as reminders of the guise. Items were evenly split between male and female voices. The guises were labeled as "Language Variety 1" and "Language Variety 2," with no other information provided about the varieties beyond what participants heard in the audio files.

In the **rating phase**, participants rated sentences on a seven-point Likert-type scale based on how likely a speaker of a given variety would be to produce them. The exact prompt that appeared on each page was as follows:

> Try to imagine whether someone who speaks Language Variety {1/2} would say each sentence. Provide your rating for how likely they would be to say it that way. Remember, focus on how the sentence is put together, not on what the sentence is about. We want to know how people talk in real life, not the grammar rules that are sometimes taught in school.

The levels of the scale were labeled as follows: "1-Very unlikely," "2-Unlikely," "3-Somewhat unlikely," "4-Neutral," "5-Somewhat likely," "6-Likely," "7-Very likely." Unlike the introduction phase, the sentences in the rating phase were presented in written form, not audio form. Participants rated a total of 52 sentences per guise, with one additional "catch" trial per guise (e.g., "Select option three to show you read this closely"). Sentences were presented in a fixed, pseudorandom order, with no consecutive items of the same grammatical phenomenon (see below).

There were 17-18 items on the page at given time, with reminder "introduction phase" items between each page.

At the end of the session, participants completed a demographic and linguistic background questionnaire. In addition to questions about age, race, gender, and languages spoken, participants provided ZIP codes for and described the community where they grew up and where they currently reside. They also provided information on the regional, educational, and occupational backgrounds of their primary caregivers. All questions included the option for open-ended responses.

2.1.2 Materials

The items in the rating phase involved three grammatical phenomena: regularized SVA with verbal (S) (21), regularized SVA with (WAS) (22), and the null copula (23). Each item appeared in one of four conditions,[3] in approximate order of predicted acceptability: grammatical in both varieties ("Grammatical-Both"), grammatical in AAL only ("Grammatical-AAL"), verb morphology that is ungrammatical in both varieties ("Ungrammatical-Agreement"), and another error outside of verb morphology ("Ungrammatical-Other"). All sentences were adapted from utterances found in CORAAL. A full list of sentences can be found in Appendix A.

(21) a. **Grammatical-Both:** They like to play a lot.

b. **Grammatical-AAL:** She like to play a lot.

c. **Ungrammatical-Agreement:** *They likes to play a lot.

d. **Ungrammatical-Other:** *Them like to play a lot.

(22) a. **Grammatical-Both:** I was crying.

[3] An additional exploratory condition was included for null copula items but is excluded here; this is described and analyzed for different participant groups in Chapter 3.

b. **Grammatical-AAL:** They was crying.

c. **Ungrammatical-Agreement:** *I were crying.

d. **Ungrammatical-Other:** *I was cry.

(23) a. **Grammatical-Both:** I'm going crazy right now.

b. **Grammatical-AAL:** He going crazy right now.

c. **Ungrammatical-Agreement:** *I going crazy right now.

d. **Ungrammatical-Other:** *He'll going crazy right now.

2.1.3 Results

Results are plotted in Figure 2.2. As we can see, participants gave high likelihood ratings to sentences that are grammatical in both varieties, and lower likelihood ratings to sentences that are ungrammatical in both varieties. In other words, participants rated grammatical sentences as likely and ungrammatical sentences as unlikely for both MAE and AAL. There was also a clear difference between guises in the Grammatical-AAL condition, where participants gave high likelihood ratings in the AAL guise, but low likelihood ratings in the MAE guise, indicating that sentences like *They was crying* (22b) are much more likely in AAL than in MAE. There was also an apparent effect of guise order, with more variability in the data when MAE is presented first, and participants gave higher likelihood ratings in to Grammatical-AAL items in the MAE guise when the MAE guise was presented first.

While there were many possible comparisons, given the small pilot data set, statistical analysis focused on the difference between guises in the two conditions where verb morphology is different from MAE: Grammatical-AAL and Ungrammatical-Agreement. Thus, the 2x2 (guise by grammaticality condition) comparison was evaluated statistically using a linear mixed-effects model, where response (centered at 0) was the dependent variable. There were fixed effects of

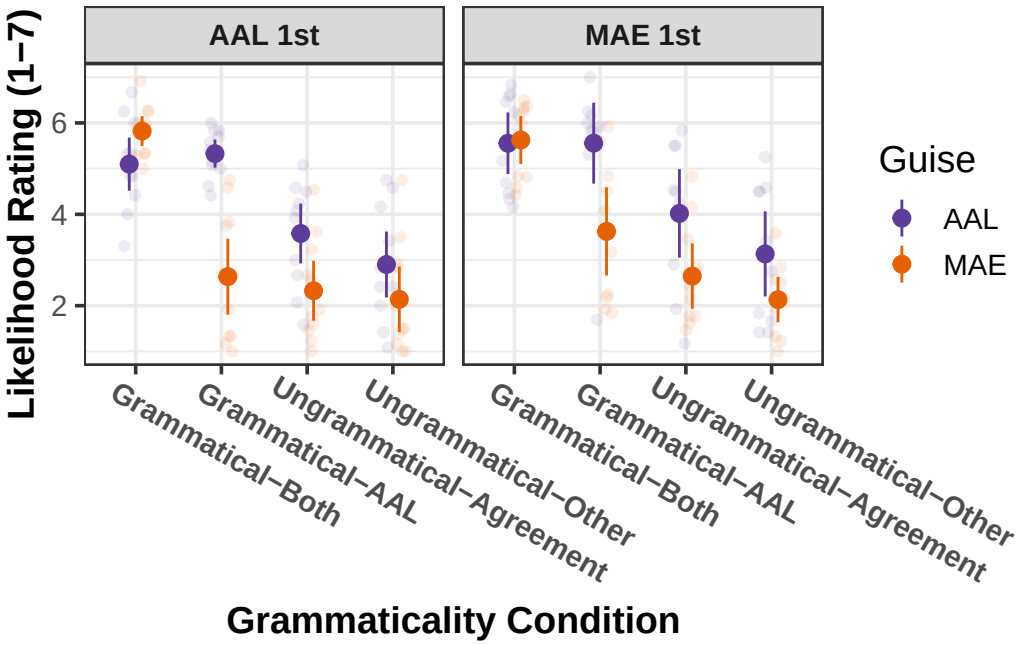

Figure 2.2: Experiment 1 results. Participant ratings are provided for each guise and grammaticality condition, separated by first guise presented. Semi-transparent points represent participant-level means, and error bars represent 95% confidence intervals, calculated as the standard error of participant-level means multiplied by 1.96.

guise and grammaticality condition (effects coded, with the AAL guise and Grammatical-AAL conditions coded as -0.5, and the MAE guise and Ungrammatical-Agreement conditions coded as 0.5), random intercepts by participant and item, and random slopes by guise and condition for both participant and item. All statistical analyses were completed in R version 4.2.2 (2022-10-31) (R Development Core Team, 2011), using the `lme4` package (Bates et al., 2015) version 1.1.31; p values were estimated using Satterthwaite's method for degrees of freedom, as implemented by the `lmerTest` package (Kuznetsova et al., 2017) version 1.1.31. Regression tables for all models can be found in Appendix A.

There was a significant effect of grammaticality condition ($\hat{\beta}^{\star} = -0.27$, $p < 0.001$), meaning that participants gave significantly lower ratings for the Ungrammatical-Agreement condition, compared to the Grammatical-AAL condition; there was a significant effect of guise ($\hat{\beta}^{\star} = -0.43$, $p < 0.001$), meaning that participants gave significantly lower ratings in the MAE guise, compared

to the AAL guise; and there was a significant interaction ($\hat{\beta}^{\star} = 0.13$, $p < 0.001$), indicating that the difference between conditions was greater in the AAL guise, compared to the MAE guise.

2.1.4 Interim discussion: Experiment 1

This experiment provided clear evidence that many MAE-dominant listeners recognize regularized SVA and attribute it to AAL. The evidence regarding the granularity of this knowledge was somewhat mixed. On the one hand, in the AAL guise, participants gave higher ratings to Grammatical-AAL items than to Ungrammatical-Agreement items, indicating that their knowledge is more granular than "verb morphology in AAL is different from MAE." On the other hand, there was a baseline guise effect, where even larger phrase structure errors were rated as more likely in AAL, compared to MAE. This might reflect familiarity with the specific grammatical rules of AAL, combined with the more general expectation that AAL grammar differs from that of MAE.

Regardless, this new method proved to be an effective tool for measuring participants' knowledge of different varieties, despite the complex metalinguistic demands. The apparent order effect is an important consideration for future implementation of this task. It makes sense that the contrast between guises would be clearer to participants when AAL is presented first. When AAL is presented first, the presence of non-mainstream grammar and the racialization are clear to most listeners. Then, when MAE examples are played in the instruction phase for the second guise, they are interpreted in comparison to AAL. However, when MAE is presented first, participants might be less certain what the comparison should be, and they might notice some regional phonetic features that are marked to them, depending on their background. Thus, when MAE is presented first, participants could be more likely to interpret the guise as including non-mainstream White varieties. This could be an interesting area for future research, but for my research questions, it

probably means that AAL should always be presented first, rather than counterbalancing across participants.

2.2 Experiment 2A: Transcription Task

The rating task in Experiment 1 showed that participants have some knowledge of regularized SVA. In terms of the Labov (1973) framework, they *recognized* it and *evaluated* that it is associated with AAL. However, in the rating task, participants were always confronted with the form of interest, so results from that task do not tell us what form listeners would *predict* in a neutral context. To determine such *prediction* knowledge, a transcription task inspired by the sociolinguistic repetition task (Buson et al., 2018; Engel and Hanulíková, 2020) was used. My version more closely aligns with the Engel and Hanulíková (2020) version of the task, since stimuli were ambiguous; the Buson et al. (2018) approach involved stimuli that mixed sociolinguistic styles and added an intervening task between audio presentation and transcription.

2.2.1 Methods

In this experiment, MAE-speaking participants listened to a sentence and transcribed what they heard. Sentences were recorded by either a speaker of AAL or MAE. As described below, a critical portion of the sentence was acoustically ambiguous, meaning that participants needed to rely on their knowledge of each speaker's grammar to resolve the ambiguity.

2.2.1.1 Materials

Speakers. All items were recorded by one White female speaker of MAE from New Jersey in her twenties and one Black female speaker of AAL from Maryland in her thirties. The Black female speaker also speaks MAE, but she only recorded AAL items due to findings by Weissler

(2021) of complex interactions between dialect and the racialization of the speaker. These two voices will be described as "guises." As described below, each speaker only recorded items that would be grammatical in the dialect they were representing, and the items were recorded in a random order so that the dialect-establishing filler items were present throughout each recording session.

Critical items. Critical items targeted two SVA phenomena present in AAL: zero-marked third-person singular on simple present tense verbs (S) (24) and *was*-leveling (WAS) (25). All critical items were adapted from naturally-occurring sentences in CORAAL. Adaptations included the removal of references to specific places, people, and events; creation of complete sentences from partial or disfluent clauses; and removal of additional contrastive grammatical features. In the case of (WAS) items, phrases such as *back then* were added in order to bias participants toward past tense; otherwise, a present-tense copula might have been compatible with the context. Items were designed so that the presence (S) or (WAS) would be ambiguous in the acoustic signal. In the case of (S), this was accomplished by using only verbs that ended in a voiceless stop, followed by a word that begins with /s/; this means that the two /s/ segments would be coarticulated, making (24b) indistinguishable from (24c). In the case of (WAS), this was accomplished through white noise masking.

(24) (S) Example

a. **Original CORAAL item:** . . . which I go visit sometimes, I- I still do

b. **AAL target item:** She visit sometimes.

c. **MAE recording version:** She visits sometimes.

(25) (WAS) Example

a. **Original CORAAL item:** That was during World War II when things was pretty

hot.

b. **AAL target item:** Things was pretty hot during the war.

c. **MAE recording version:** Things were pretty hot during the war.

For all of these items, the dependent variable is whether participants transcribe MAE SVA (24c, 25c) or regularized SVA (24b, 25b). To ensure that the items were equivalent at the ambiguous region and able to be produced naturally by each speaker, all critical items were spliced from two sentences, exemplified below with the included portion in brackets. The first portion went through the subject of the sentence, which was third person singular for (S) (26a, 26b) and plural for (WAS) (27a, 27b). The second portion started at the verb and contained the remainder of the sentence, preceded by a plural subject for (S) (26c) and a singular subject for (WAS) (27c). Crucially, these second portions are zero-marked for (S) or use *was* regardless of dialect, meaning that the acoustic information with regard to SVA was equivalent between the two guises. This also means that the acoustic signal was biased toward regularized SVA. During recording sessions, each speaker recorded both versions in sequence. Speakers were told to maintain the same prosody for both versions to facilitate future splicing.[4]

(26) a. Part 1, AAL: [She] visit sometimes.

b. Part 1, MAE: [She] visits sometimes.

c. Part 2: They [visit sometimes].

(27) a. Part 1, AAL: [Things] was pretty hot during the war.

b. Part 1, MAE: [Things] were pretty hot during the war.

c. Part 2: Everything [was pretty hot during the war].

[4]Since I hope that this method is used more in the future, I should note that this is an important step, but it is not trivial; speakers do not naturally deliver two minimally different sentences in sequence with identical prosody. I had to listen carefully during the recording session and give extensive feedback to the speakers. This care was worthwhile, since the splices proved undetectable.

To the extent possible, obstruents were used at the splice points to facilitate easy identification of word boundaries; for (S) items, all verbs began with an obstruent, and for (WAS) items, the word after the auxiliary began with an obstruent other than /s/, to avoid coarticulation with the end of *was*. (Since pronominal subjects, such as *they* and *we*, were sometimes used, it was not always possible to have an obstruent at splice points for the (WAS) items.) For (WAS) items, the word *was* was replaced by white noise, leaving the first 50 ms of the /w/ to increase the likelihood that participants would identify the masked word as *was* or *were*. All splice points, including the start and end of *was* tokens, were identified manually using Praat, and splicing was accomplished using a Praat script, which automatically moved each splice point to the nearest zero crossing. Prior to the norming process described below, all spliced items were reviewed to confirm that they sounded natural (other than the presence of white noise).

Filler Items. In addition to the racialization of the speakers, filler items were used to establish each guise. These items primarily used stereotyped features of AAL: null copula (28) and negative concord, both without (29) and with (30) negative inversion. Items with negative inversion were adapted from CORAAL, while the non-inverted negative concord and null copula items were designed for future comparison with visual world paradigm studies, requiring more tightly-defined sentence templates. The MAE guise used MAE-equivalent versions of each sentence: single negation, non-inverted sentences, and overt copulas. The use of variable SVA phenomena was minimized in these items to ensure that participants were relying on their pre-existing knowledge of SVA variation. Thus, for negative concord items, *don't* never appeared with a singular subject, and plural subjects were avoided entirely for the null copula items to avoid introducing the *is*/*are* distinction in the MAE filler items. Some of the AAL negative concord items used *ain't* where MAE uses *didn't*; this is common in AAL, and the speaker confirmed that it sounded most natural to her. Variables such as (ING) do appear in these items and were not

deliberately manipulated; speakers recorded them according to what felt natural in the recording session.

(28) Null copula

a. **AAL:** She watching the clock.

b. **MAE:** She's watching the clock.

(29) Negative concord (no inversion)

a. **AAL:** He ain't read no book.

b. **MAE:** He didn't read any book.

(30) Negative inversion

a. **AAL:** Can't nobody stop you from trying.

b. **MAE:** Nobody can stop you from trying.

Since white noise was present in half of the critical items—the (WAS) items—it was also used in half of the filler items. White noise was placed in a variety of places, always over the vowel of a stressed syllable. This was to ensure that the intended word was easy to recover, making the task of "filling in" *was* or *were* more natural in the critical items.

Item Norming. All items were normed with a separate group of participants ($n = 29$) on Prolific. Participants completed a shortened version of the task in Qualtrics, where they transcribed a random subset of 50 items (30 AAL, 20 MAE, since a larger pool of items was being tested for AAL). Each item was presented about 4-5 times per guise, and items were blocked by guise, with the order of guises counterbalanced across participants. Participants were paid $1.75, and the median completion time for the task was about 12-13 minutes.

Items were removed from the item pool if there was generally low accuracy, or if a given mis-transcription was made more than once. This was most important when such choices affected

the transcription of SVA. In the case of (S) items, participants preferred to transcribe the verb in the past tense for regular verbs ending in /p/ or /k/ (31). This makes sense, since the past tense would produce a /pt/ or /kt/ cluster, both of which are commonly reduced in both varieties, particularly AAL. While this poses a challenge for the splicing design, which requires verbs that end in a voiceless stop, two types of items avoided this problem: irregular verbs (e.g., *get*, *keep*), where the past tense has more differences from the present tense, and verbs ending in /t/ (e.g., *create*), where past tense would require epenthesis.

(31) a. **Intended Sentence:** The boy dip(s) stuff in the water.

b. **Common Transcription:** The boy dipp**ed** stuff in the water.

For (WAS) items, the primary concern was whether participants would transcribe *was* or *were* at the white noise, rather than some other word that might yield a complete sentence. There was only an issue for one item, which was excluded due to a preference for the word *went*:

(32) a. **Intended Sentence:** On Friday, they **was** down there.

b. **Common Transcription:** On Friday, they **went** down there.

Based on this norming process, the final item pool consisted of the following: 18 (S) items, 18 (WAS) items, 12 negative concord items, 12 negative inversion items, and 12 null copula items. All items were equalized in intensity to 68 dB using Praat. The full final item list can be found in Appendix A.

2.2.1.2 Participants

Participants ($n = 60$) were recruited from Prolific, using the same demographic filters as Experiment 1. (These were different individuals from Experiment 1; participants were not permitted to complete more than one of the studies presented in this chapter.) They were initially

paid $3.50, with the intention of paying $10.50 per hour for a 20-minute task. However, the actual completion time was highly variable, ranging from about ten minutes to about one hour, with a median completion time of about 26 minutes. Because of this, I awarded a bonus payment of $1.14 to all participants who completed the study ($n = 57$), re-establishing the hourly rate of $10.50. The variable completion time likely represents the combination of differences in typing speed and variable break times, as participants could take a break whenever they wanted to do so. This sample was diverse in terms of age (mean=35.8, s.d.=9.4), gender identity (32 male, 25 female), and region of the U.S (see summary by state in Table 2.3). Most had completed a four-year college degree or more (Table 2.4).

Table 2.3: Number of participants who grew up or currently live in each U.S. state for Experiment 2A.

State	Childhood	Current
Illinois	8	4
Florida	6	9
New Jersey	5	3
California	4	3
New York	4	2
Massachusetts	3	2
Pennsylvania	2	4
Tennessee	2	3
Indiana	2	2
Louisiana	2	2
Wisconsin	2	2
Michigan	2	1
Kentucky	1	2
Maryland	1	2
Mississippi	1	2
North Carolina	1	2
South Carolina	1	2
Colorado	1	1
Connecticut	1	1
New Mexico	1	1
Ohio	1	1
Texas	1	1
Utah	1	1
Virginia	1	1

State	Childhood	Current
Alabama	1	0
Georgia	1	0
Minnesota	1	0
Missouri	0	1
Nevada	0	1
New Hampshire	0	1

Table 2.4: Education levels, Experiment 2A

Level of Education	Number of Participants
Graduate School	13
4-year college degree	29
Associate's/Technical degree, trade school, or some college	8
High school diploma	6
GED	1

Participants' self-reported usage of and exposure to each dialect is provided in Figure 2.3. As we can see, participants overwhelmingly selected the lowest level of the scale for AAL usage and the highest level of the scale for MAE usage. For exposure, they primarily selected the highest level of the scale for MAE and selected the center of the scale for AAL. These ratings generally confirm that the demographic factors successfully recruited a monodialectal sample of MAE speakers.

2.2.1.3 Procedure

Participants completed the study in Qualtrics on their own devices. After completing the consent form, they read the following instructions:

> In this study, you will be listening to sentences that represent two different language varieties. Your job is to listen to each sentence and type what you hear. Make sure you type each word. Sometimes, there might be issues with the audio file. If this happens, just do your best to type what you think the person said.

They then answered a comprehension question to confirm that they understood the task:

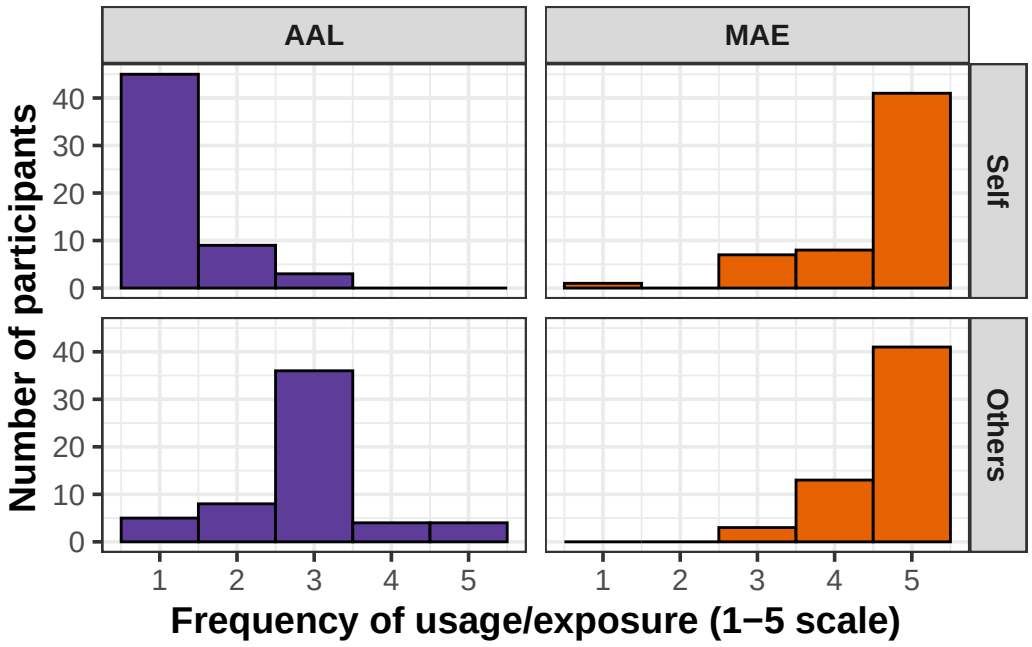

Figure 2.3: Self-reported usage of and exposure to AAL and MAE for participants in Experiment 2A

What are you supposed to do in this study?
A) Type each word of the sentences I hear [Correct answer]
B) Edit each sentence to make it better
C) Decide which sentence makes the most sense

These instructions were intended to make it clear that the guises represented different varieties and to make sure that participants did not attempt to standardize sentences or simply omit words that were masked by white noise.

After this, they completed the sentence transcription task. Items were blocked by guise, with the order of guises counterbalanced across participants. The guises were introduced as "Language Variety 1" and "Language Variety 2," respectively. To immediately establish the guise, at the beginning of each block, three filler items appeared (one per type, with one item containing white noise), followed by the remaining 33 items (18 critical items) in random order. Critical items were drawn from one of three lists for each guise, yielding six possible list pairings. Each list was presented equally using the "evenly present elements" option in Qualtrics. One item was

presented on a page without the option to return to the previous item, and each audio file could be played as many times as the participant wanted.

After the sentence transcription task, participants answered questions about their linguistic and demographic backgrounds. They first answered an open-ended question—"What did you notice about the different types of language used in the experiment?"—then answered specific questions about each guise. They were asked to rate on a scale from 1-5 how often they use each language variety and how often they interact with users of each variety, followed by an open-ended question about their experience with each variety. An audio file exhibiting the (WAS) variable, without masking, was used to remind the participants of each guise and make subject-verb agreement salient in their selections regarding linguistic experience. After this, they completed a demographic questionnaire, followed by questions about the area where they grew up, the area where they currently live, and the backgrounds of their primary caregivers during their childhood.

2.2.2 Results

For (S) items, responses were coded as "AAL" if they contained the subject (in singular form) and the zero-marked target verb, and they were coded as "MAE" if the verb was *s*-marked. For (WAS) items, responses were coded as "AAL" if they contained the correct subject followed by *was*, and they were coded as "MAE" if the subject was followed by *were*. Coding was accomplished using string-matching expressions in R; punctuation and capitalization were not taken into account. Any responses that did not match these criteria were treated as missing data (12.79% of items in the AAL guise, 7.35% of items in the MAE guise).

Results are plotted in Figure 2.4. As we can see, participants showed an overall preference for using MAE SVA, generally responding with regularized SVA less then half the time, except for (WAS) in the AAL guise. However, they were more likely to transcribe regularized SVA in

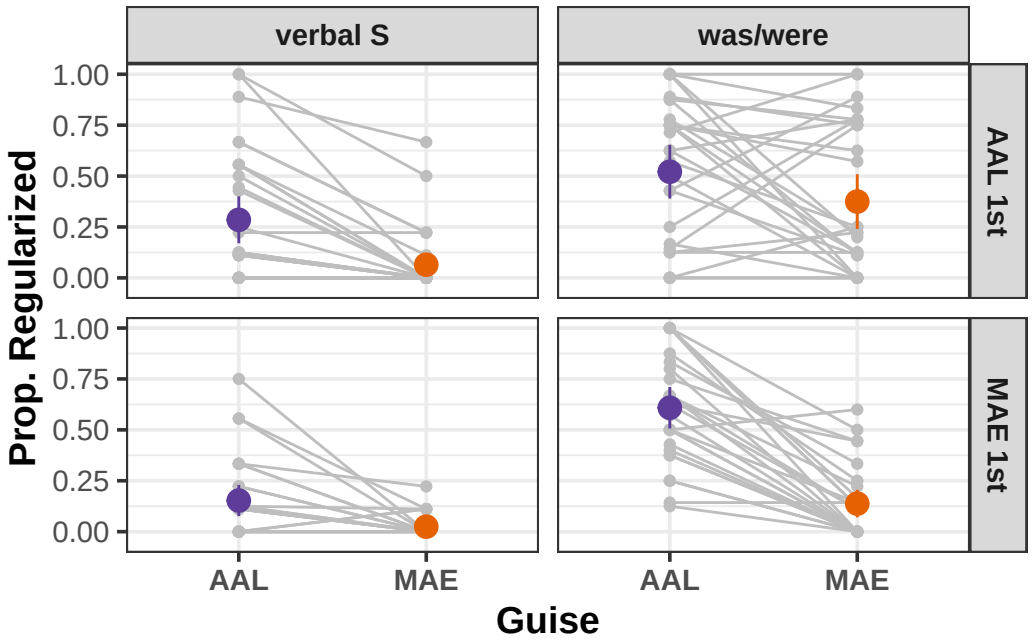

Figure 2.4: Results for Experiment 2A. Regularized responses are provided as a proportion of scorable items, split by SVA phenomenon and first guise presented. Gray lines and points represent participant-level proportions, with grand means provided in color. Error bars represent 95% confidence intervals, calculated based on participant-level means. Note that a given line might represent more than one participant, particularly in the case of participants who never made a regularized transcription.

the AAL guise than in the MAE guise, and they were more likely to transcribe *was* with a plural subject than zero-marked (S). There was also an apparent effect of guise order, particularly for (WAS); participants were more likely to transcribe *was* for the MAE guise if they had already heard AAL.

These visual results were confirmed using a logistic mixed-effects model, where the dependent variable was whether a participant transcribed using regularized SVA. There were fixed effects of guise, SVA phenomenon, and the guise-by-phenomenon interaction; participant- and item-level random intercepts; and guise-by-participant and guise-by-item random slopes. All fixed-effects terms were effects-coded to allow for interpretation as main effects (-0.5 for the AAL guise and (S), 0.5 for the MAE guise and (WAS)). There was a significant intercept ($\hat{\beta}^{\star} = -2.41$, $p < 0.001$), indicating that participants were less likely to transcribe using regularized SVA, relative to MAE

variants; a significant effect of guise ($\hat{\beta}^{\star} = -1.63$, $p < 0.001$), indicating that participants were less likely to transcribe regularized SVA in the MAE guise, compared to the AAL guise; and a significant effect of SVA phenomenon ($\hat{\beta}^{\star} = 1.51$, $p < 0.001$), indicating that participants were more likely to transcribe regularized SVA for (WAS) than (S), regardless of guise. There was no significant phenomenon-by-guise interaction ($\hat{\beta}^{\star} = 0.26$, $p = 0.12$).

The addition of guise order, along with all possible two- and three-way interactions, improved data-model fit ($X^2(4) = 21.73$, $p < 0.001$). To assist interpretation of the order effects, this model was re-fitted with dummy-coded fixed effects, with the AAL guise, (S), and AAL-first order as the reference levels. There was a significant intercept ($\hat{\beta}^{\star} = -1.53$, $p < 0.001$), indicating that participants provided regularized transcriptions less than half the time for (S) items in the AAL guise, when this guise was presented first. There was a significant effect of guise ($\hat{\beta}^{\star} = -3.55$, $p < 0.001$), indicating that participants provided regularized transcriptions less often for (S) items in the MAE guise, compared to the AAL guise, when the AAL guise was presented first. There was a significant effect of SVA phenomenon ($\hat{\beta}^{\star} = 2.01$, $p < 0.001$), indicating that participants provided regularized transcriptions more often for (WAS) items than (S) in the MAE guise, when the AAL guise was presented first. There was also a significant guise-by-phenomenon interaction ($\hat{\beta}^{\star} = 2.07$, $p = 0.007$), indicating that when the AAL guise was presented first, regularized transcriptions were more common in the MAE guise for (WAS) than for (S).

Turning to effects of guise order, there was a marginal effect of guise order ($\hat{\beta}^{\star} = -0.83$, $p = 0.07$), indicating a possible decrease in regularized transcriptions for (S) in the AAL guise, when the AAL guise was presented second, rather than first. There was a significant interaction between guise order and item type ($\hat{\beta}^{\star} = 1.05$, $p = 0.008$), indicating higher rates of regularized transcriptions in the AAL guise for (WAS) when MAE was presented first, compared to when AAL was presented first. There was also a significant three-way interaction ($\hat{\beta}^{\star} = -2.73$, $p <$

0.001), indicating a decreased likelihood of regularized transcriptions in the MAE guise for (WAS) when MAE was presented first, compared to when MAE was presented second. There was no significant interaction between guise and first guise ($\hat{\beta}^{\star} = 0.1$, $p = 0.91$), indicating no difference in regularized transcriptions for (S) in the MAE guise, regardless of the order of guise presentation.

2.2.3 Interim discussion: Experiment 2A

This experiment provided evidence that most monodialectal MAE speakers have some knowledge of SVA variation and attribute the regularized variants to AAL. They demonstrate this knowledge in neutral contexts, where they could interpret the stimuli according to their own dialect, indicating a greater level of knowledge than merely recognizing the variants and identifying the indexical value of each variant. In addition to demonstrating knowledge of SVA, this experiment provides some support for the idea that SVA-based variation plays a top-down role in guiding perception. Of course, the transcription task is an imperfect measure of perception, since participants could perceive the ambiguity and then consciously decide which variant to used based on their stereotypes of each dialect, but only a handful of participants made any mention of the manipulation in the open-ended question.

This overall finding comes with multiple caveats, some of which might reflect areas where participants' knowledge of variation is imperfect. First, there was a substantial difference between the two SVA phenomena under examination, where participants were much less likely to transcribe the regularized (S) variant (*She visit sometimes*) than the regularized (WAS) variant (*Things was pretty hot*). This might reflect different levels of exposure to the two forms or some underlying difference between free and bound morphemes in their availability to social evaluation, along the lines of the Interface Principle (Labov et al., 1993). This difference might also be one of salience, since the (S) variable, in addition to being bound, is one phone and could be coarticulated in the

stimuli, while the (WAS) variable, in addition to being free, is multiple phones and was made salient by the presence of white noise. These explanations are not mutually exclusive.

Second, participants still showed an overall preference for MAE forms. This might reflect a lingering bias to transcribe according to their primary dialect, reflecting some competition between their own grammar and their representation of other grammars, and it could also reflect a general bias toward using MAE variants when engaged in writing in the context of a university-based study. Third, participants sometimes transcribed the regularized variant in the MAE guise—particularly for (WAS) when the AAL block was presented before the MAE block. One reason for this is the way the materials were constructed, which deliberately biased acoustic cues toward regularized SVA. Because of this, participants were actually hearing a token of the AAL variant, and it is possible that subtle acoustic cues, including the first 50 ms of *was*, reinforced this point. Thus, when participants listened to the MAE guise, they were forced to resolve potential competition between their mental model of the speaker's grammar, which favored MAE variants, and the acoustic cues, which favored regularized variants. Future work could test this by presenting the original, non-spliced items, and I partially address this in Experiment 2C, where only 20 ms from the onset of *was* is included.

However, this acoustic difference does not explain the effect of guise order, where participants were more likely to transcribe *was* in the MAE guise when it followed the AAL guise. I speculate here that the presence of (WAS) items in the AAL guise increased the availability of regularized *was* in this context, and when subtle acoustic cues favored *was*, participants were more likely to transcribe this regularized variant. Though this effect might be limited to the task, it is also possible that this finding is of theoretical interest. Participants likely began with the assumption that the experiment is an MAE-dominant context. The AAL guise provided evidence to the contrary, so participants expected common non-mainstream variants to be available.

While this task is far removed from an ecologically valid discourse, these order effects raise the question of the unit of analysis when listeners adjust their grammatical expectations. One possibility is that listeners form expectations for an individual speaker. Thus, in a discourse with both MAE and AAL-speaking participants, this listener would expect regularized SVA from AAL speakers and MAE SVA from MAE speakers. Another possibility is that listeners form expectations for a discourse, expecting only MAE, only AAL, or both, regardless of who is speaking. This could partially explain guise order effects: when the MAE speaker followed an AAL speaker, the option for regularized SVA remained present in the discourse. An advantage to such a strategy is that it requires less of the listener; they must only track the varieties available in the discourse without also mapping each option to a given speaker. Additionally, it is plausible to expect other participants in a discourse to accommodate each other (Giles and Ogay, 2007), so regularized SVA is more likely from Speaker B after it has already been used by Speaker A.

More broadly, we might ask the following: what generalizations were participants drawing about each guise? In this guise design, multiple factors are confounded: the intended dialect, individual speaker identity, racialization of the speaker, minoritization of the speaker, in-group vs. out-group status of the speaker, sociophonetic cues in each item, and morphosyntactic cues in the filler items. While AAL might be the most prominent variety where verbal (S) is absent and (WAS) is leveled, this can also be found in Southern White English and many other varieties. Thus, it is possible that participants thought of the AAL guise as a guise with "common errors," rather than AAL in particular. Such differences could be manipulated experimentally as a test of the granularity of each listener's social evaluation. Overall, however, we have seen that listeners do adjust their grammatical expectations depending on the speaker, at times even expecting variants that they do not produce. However, listeners vary in the degree to which they do this.

2.3 Experiment 2B: Mixed Design

The block order effects in Experiment 2A present both theoretical and methodological questions. As I noted above, on a theoretical level, we might ask the degree to which listeners switch between grammars depending on their interlocutor, or broadly consider the available grammatical options, regardless of the speaker. On a methodological level, it is useful to know how robust the observed effects are to different decisions about stimulus order. Since one goal is to develop a task to compare different groups (see Chapter 4), it would be preferable to have a version where it is not necessary to counterbalance order between participants. To address these questions, Experiment 2B replicated Experiment 2A, but with one change: instead of presenting AAL and MAE in separate blocks, they were mixed together, with items appearing in random order. As in Experiment 2A, the first three items from each voice were guise-establishing filler items.

2.3.1 Participants

Participants ($n = 60$) were recruited from Prolific and were paid $4.64 (the adjusted payment for Experiment 2A). Of these participants, 56 identified as "White, non-Hispanic" and no other race and were included in analysis.[5] This sample was diverse in terms of age (mean=34.6, s.d.=9), gender identity (33 male, 22 female), and region of the U.S (see summary by state in Table 2.5). Participants' level of education is provided in Table 2.6; a majority of participants had a post-secondary education.

[5] All participants identified as "White, non-Hispanic" for the other experiments reported in this chapter as well, as this aligned with the pre-screening criteria; it is unclear why a few participants with other racial/ethnic identities responded here, perhaps due to subtle differences between Prolific's pre-screener and my demographic questions.

Table 2.5: Number of Experiment 2B participants who grew up or currently live in each U.S. state.

State	Childhood	Current
New York	6	6
Illinois	5	4
California	5	3
New Jersey	5	2
Texas	4	7
Arizona	4	4
Maryland	3	3
Massachusetts	3	1
Florida	2	3
Ohio	2	2
Pennsylvania	2	2
Georgia	2	1
Maine	2	1
Wisconsin	2	1
Virginia	1	3
Colorado	1	2
Michigan	1	2
Iowa	1	1
Kentucky	1	1
Minnesota	1	1
Tennessee	1	1
Utah	1	0
Washington	1	0
Hawaii	0	1
North Carolina	0	1
Oklahoma	0	1
Oregon	0	1
South Carolina	0	1

Table 2.6: Education levels, Experiment 2B

Level of Education	Number of Participants
Graduate School	11
4-year college degree	26
Associate's/Technical degree, trade school, or some college	9
High school diploma	9
Less than high school diploma	1

Participants' self-reported usage of and exposure to each dialect is provided in Figure 2.5.

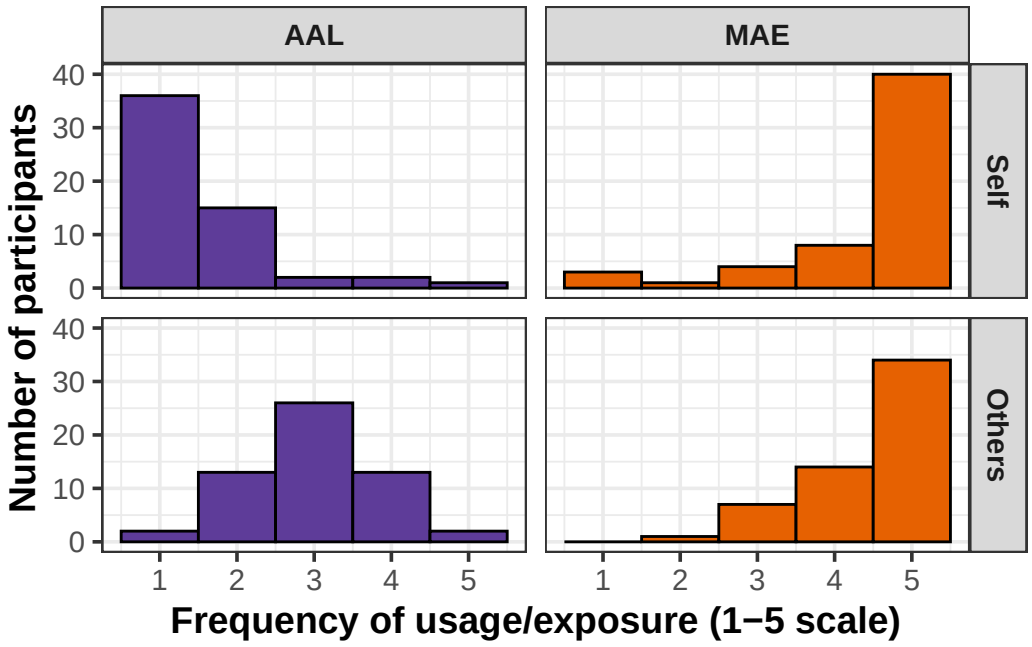

Figure 2.5: Self-reported usage of and exposure to AAL and MAE, Experiment 2B participants

These largely mirror the self-report data observed above, with slightly elevated responses to the 2/5 category for AAL usage.

2.3.2 Results

The same string-match procedure was used to code the data. Any responses that did not match these criteria were treated as missing data (13.04% of items in the AAL guise, 7.38% of items in the MAE guise). Results are plotted in Figure 2.6. The results are qualitatively the same as Experiment 2A: participants more frequently provieded regularized responses in the AAL guise, compared to the MAE guise, and they more frequently provided regularized responses for (WAS) items, compared to (S) items.

These visual results were confirmed using the same logistic mixed-effects model specification as above: fixed effects of guise, SVA phenomenon, and guise-by-phenomenon interaction; participant- and item-level random intercepts; and guise-by-participant and guise-by-item random slopes. All fixed effects terms were effects-coded. Inferences from this model are identical.

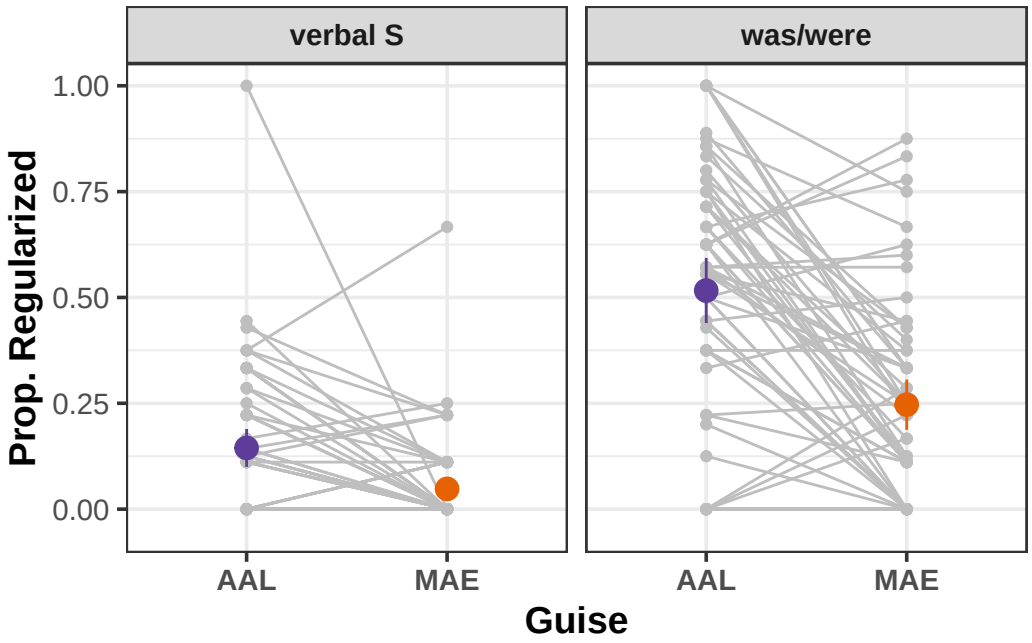

Figure 2.6: Results for Experiment 2B. Regularized responses are provided as a proportion of scorable items, split by SVA phenomenon. Gray lines and points represent participant-level proportions, with grand means provided in color. Error bars represent 95% confidence intervals, calculated based on participant-level means. Note that a given line might represent more than one participant, particularly in the case of participants who never made a regularized transcription.

There was a significant intercept ($\hat{\beta}^{\star} = -2.28$, $p < 0.001$), indicating that participants were less likely to transcribe using regularized SVA, relative to MAE variants; a significant effect of guise ($\hat{\beta}^{\star} = -1.08$, $p < 0.001$), indicating that participants were less likely to transcribe regularized SVA in the MAE guise, compared to the AAL guise; and a significant effect of SVA phenomenon ($\hat{\beta}^{\star} = 1.31$, $p < 0.001$), indicating that participants were more likely to transcribe regularized SVA for (WAS) than (S), regardless of guise. There was no significant phenomenon-by-guise interaction ($\hat{\beta}^{\star} = 0.03$, $p = 0.86$).

2.3.3 Interim discussion: Experiment 2B

This version replicated the results from Experiment 2A; participants were more likely to provide regularized responses in the AAL guise, compared to the MAE guise, and they were more likely to transcribe the regularized variant for (WAS) than (S). This suggests that even in a highly

variable context, listeners adapt their grammatical expectations depending on their interlocutor's grammar. This also means that for the purpose of comparing between-participants factors, we can use this version of the task without adding complicating order effects.

2.4 Experiment 2C: Multiple voices

In Experiments 2A and 2B, there were only two voices, each corresponding to one guise. This weakens the inference that the observed effects were about dialect differences, rather than some idiosyncratic speaker factor. To begin to address this issue, I introduced two additional voices, both male college students in their early twenties from Maryland. One speaker represented MAE, and the other represented AAL.

In addition to this change, I made some changes to the items. First, I reduced the duration of the /w/ token in the (WAS) items from 50 ms to 20 ms, with the remaining 30 ms replaced by white noise. This reduced participants' ability to rely on the partial word to guide their transcription. Second, additional dialect-neutral versions of the SVA items (26c, 27c) were included as fillers, amounting to half of the number of critical items (due to a scripting error, only female voices were used for this item type). This means that the presence of ambiguous SVA would not always interact with guise. Third, the null copula filler items were removed and replaced with additional negation fillers, in anticipation of using this task before the rating task of Experiment 1 in the same experimental session (see Chapter 3); null copula is a critical item type in the rating task, so inclusion as a filler here could alter the results of the rating task.

2.4.1 Participants

Participants ($n = 23$) were recruited from Prolific and included in analysis. Participants were paid \$4.55 for the estimated 26-minute duration; actual median was 24 minutes. This sample

represented a range of ages (mean=28.1, s.d.=8.1), gender identities (11 male, 11 female), and regions of the U.S (see summary by state in Table 2.7). Participants' level of education is provided in Table 2.8; a majority of participants had a post-secondary education.

Table 2.7: Number of Experiment 2C participants who grew up or currently live in each U.S. state.

State	Childhood	Current
Pennsylvania	3	3
Washington	3	3
California	2	3
New York	2	2
Tennessee	2	1
Indiana	1	2
Florida	1	1
Georgia	1	1
Illinois	1	1
Kentucky	1	1
North Carolina	1	1
Texas	1	1
Virginia	1	1
Louisiana	1	0
Maryland	1	0
Ohio	1	0
Colorado	0	2

Table 2.8: Education levels, Experiment 2C

Level of Education	Number of Participants
Graduate School	3
4-year college degree	10
Associate's/Technical degree, trade school, or some college	7
High school diploma	3

Participants' self-reported usage of and exposure to each dialect is provided in Figure 2.7. We can see that participants generally selected the bottom of the scale for AAL usage and the top of the scale for MAE usage. Additionally, in this version of the task, the scale was reduced to four levels to avoid a middle category; we can see the modal participant selected "2," which is

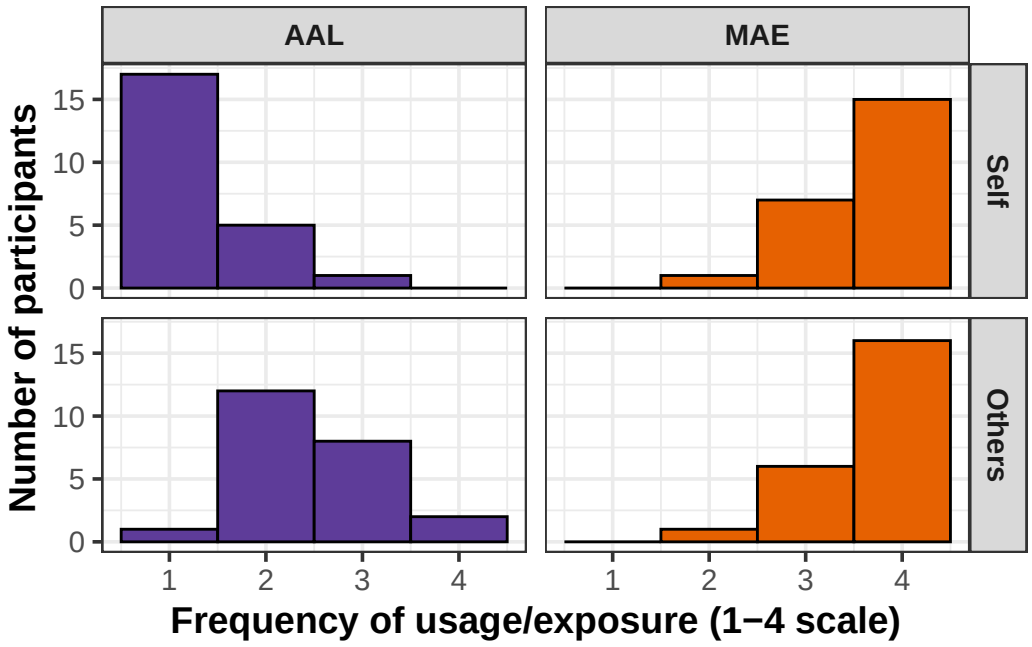

Figure 2.7: Self-reported usage of and exposure to AAL and MAE, Experiment 2C

just below the middle of the scale, for exposure to AAL.

2.4.2 Results

The same string-match procedure was used to code the data. Any responses that did not match these criteria were treated as missing data (10.07% of items in the AAL guise, 6.6% of items in the MAE guise). Results are plotted in Figure 2.8. We can see that the guise effect appears to be maintained, but the difference between SVA phenomena appears to be reduced, with a lower proportion of regularized responses for (WAS) relative to previous versions of the study.

These visual results were tested using the same logistic mixed-effects model specification as Experiments 2A and 2B. However, due to the lower number of participants, no random slopes were included to avoid issues of a singular fit. There was a significant intercept ($\hat{\beta}^{\star} = -2.94$, $p < 0.001$), indicating that participants were less likely to transcribe using regularized SVA, relative to MAE variants, and a significant effect of guise ($\hat{\beta}^{\star} = -1.05$, $p < 0.001$), indicating that participants

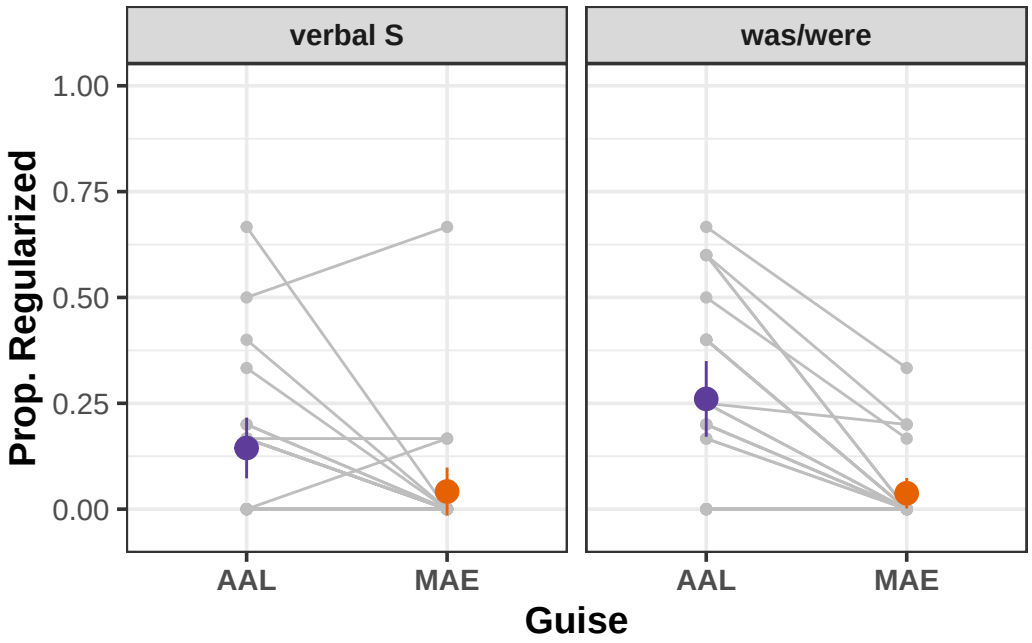

Figure 2.8: Results for Experiment 2C. Regularized responses are provided as a proportion of scorable items, split by SVA phenomenon. Gray lines and points represent participant-level proportions, with grand means provided in color. Error bars represent 95% confidence intervals, calculated based on participant-level means. Note that a given line might represent more than one participant, particularly in the case of participants who never made a regularized transcription.

were less likely to transcribe regularized SVA in the MAE guise, compared to the AAL guise. Unlike previous results, there was no significant effect of SVA phenomenon ($\hat{\beta}^{\star} = 0.17$, $p = 0.49$), and there was no significant phenomenon-by-guise interaction ($\hat{\beta}^{\star} = -0.28$, $p = 0.14$).

2.4.2.1 Guise Gender

This task introduced gender, in addition to dialect, to the guise manipulation. Responses split by gender are plotted in Figure 2.9. We can see the guise effect for both included genders and no apparent effect of gender. These visual results were tested using a logistic mixed-effects model with fixed effects of guise (dialect), gender, and their interaction, as well as participant- and item-level random intercepts. There was a significant intercept ($\hat{\beta}^{\star} = -2.98$, $p < 0.001$), indicating that participants were less likely to transcribe using regularized SVA, relative to MAE variants, and a significant effect of guise (dialect) ($\hat{\beta}^{\star} = -1.09$, $p < 0.001$), indicating that participants

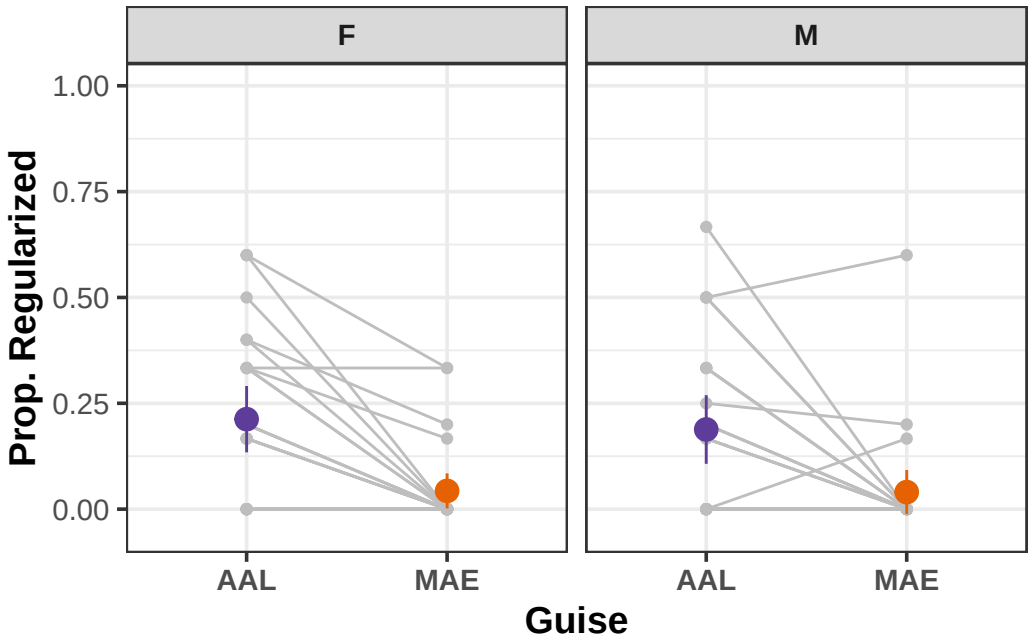

Figure 2.9: Results split by gender. Regularized responses are provided as a proportion of scorable items, split by SVA phenomenon. Gray lines and points represent participant-level proportions, with grand means provided in color. Error bars represent 95% confidence intervals, calculated based on participant-level means. Note that a given line might represent more than one participant, particularly in the case of participants who never made a regularized transcription

were less likely to transcribe regularized SVA in the MAE guise, compared to the AAL guise. There was no significant effect of gender ($\hat{\beta}^{\star} = -0.14$, $p = 0.45$), and there was no significant guise-by-gender interaction ($\hat{\beta}^{\star} = 0.02$, $p = 0.93$).

2.4.3 Interim discussion: Experiment 2C

These results show that the effects observed thus far for the transcription task are robust to the inclusion of multiple speakers per dialect. This is noteworthy for two reasons. First, since voices were presented in a random order, the task of tracking four separate voices could have been more challenging than the two voices of Experiment 2B, leading participants to have more general, experiment-wide expectations. Second, the presence of male voices could change expectations regarding standardness; if listeners expect male speakers to use more non-mainstream variants, we might have observed a general effect of gender, or an effect where the male MAE

speaker had an increased rate of regularized SVA relative to the female MAE speaker, reducing the MAE vs. AAL contrast overall.

However, the reduced difference between (WAS) and (S) items suggests that the amount of white noise masking affects participants' transcriptions. In other words, some of the higher rates of regularized SVA for (WAS) in Experiments 2A and 2B were driven by bottom-up cues favoring *was*, rather than top-down grammatical expectations. Since listeners' top-down grammatical expectations are of primary interest here, and the shorter /w/ token had no impact on missingness, it is preferable to use this version of the task moving forward.

2.5 Conclusion

Across a series of studies, I consistently observed that many speakers who primarily use MAE still have knowledge of regularized SVA and associate it with AAL. In terms of the Labov (1973) framework, I have found evidence for *recognition* of regularized SVA, *evaluation* that it is associated with AAL, and *prediction* of regularized forms in a neutral context. Additionally, the broad findings of a guise effect in the transcription task are robust to many different details of task design, suggesting that it is a useful tool for understanding listeners' perception of different sociolinguistic variables.

Despite the fact that participants showed some knowledge of regularized SVA, they were still MAE-dominant in two ways. First, in Experiment 1, they gave overall higher ratings for ungrammatical items in the AAL guise, compared to the MAE guise *even when those items were ungrammatical in AAL*. Second, in all versions of Experiment 2, participants showed a general preference for MAE SVA. Thus, it is possible that listeners with greater knowledge of AAL would perform differently on these tasks. While the present results reflect a broader sample than is common in studies that draw from the undergraduate population at one research university, they

do not allow for comparison across different levels of exposure to AAL. In Chapter 3, I will build on these paradigms and compare knowledge across different groups of listeners.

Chapter 3: Knowledge of morphosyntactic conditions in bidialectal and monodialectal listeners

In Chapter 2, we saw evidence of partial knowledge of regularized subject-verb agreement (SVA) in nationally-recruited sample of White, higher-SES participants. When presented with a regularized variant, I found that participants *recognize* the form as associated with AAL, but I found mixed evidence regarding the *granularity* of this knowledge; they were less likely to attribute regularized SVA (*They was*) than uncommon SVA (*He were*) to AAL, but they were more likely to attribute uncommon SVA to AAL than to MAE. Thus, for the less familiar variety, these participants combined granular knowledge of the difference in SVA patterns with generally relaxed expectations of grammaticality. Complementary evidence was found from the transcription task: most participants *predicted* AAL at least occasionally in a neutral context where they could still rely on the grammar of MAE to resolve an ambiguity.

All of these findings came from participants who are unlikely to use regularized SVA; based on their self-report, these participants said that they spoke MAE but not AAL, and there were no clear effects of linguistic experience in this deliberately narrow range. Given this relative lack of experience with AAL, it makes sense that they would still show a general preference for MAE forms and only partial knowledge of AAL forms. We might expect, then, that participants whose linguistic repertoire includes both varieties would show both increased granularity in their representations and increased propensity to predict different forms in a neutral context depending on the grammar of their interlocutor.

In this chapter, I ask how language experience affects linguistic knowledge, as measured by the sentence rating task and transcription task used in Chapter 2. While linguistic experience exists on a continuum, I sampled from different parts of the continuum, targeting two groups: "monodialectal" speakers of MAE and "bidialectal" speakers of both AAL and MAE.[1] Broadly, I hypothesize that bidialectals will show greater knowledge of AAL across tasks and greater propensity to adjust their linguistic expectations depending on the grammar of their interlocutor. This series of experiments was pre-registered using the Open Science Framework (osf.io/5neuy), which included pre-registration of the hypotheses, design, approach to recruitment, exclusionary criteria, and analysis plan. I note both the pre-registered plan and some changes and exploratory steps throughout this chapter.

3.1 Hypotheses

The pre-registered hypotheses anticipate a combination of factors: all participants will show some knowledge of both AAL and MAE grammatical rules, while bidialectal participants will show more knowledge of AAL grammatical rules compared to monodialectal participants. There are two ways that bidialectal participants might show increased knowledge of AAL grammar, relative to monodialectal participants. First, they might generally expect any speaker to follow the grammatical rules of AAL. This would mean that they would show granular knowledge of AAL SVA rules and predict regularized agreement, but they would not differentiate their expectations based on the grammar of their interlocutor. This idea is reflected in Hypotheses 3A and 3C. Second, they might tailor their expectations to the grammar of their interlocutor, expecting regularized SVA from AAL speakers but MAE SVA from MAE speakers. This idea is reflected in Hypotheses 3B and 3D. These hypotheses are not mutually exclusive, and it is possible that

[1] As I will discuss throughout the analysis and discussion, this simplification comes with substantial drawbacks. My usage of these terms should not be interpreted as a commitment to "monodialectalism" and "bidialectalism" as theoretical constructs.

there is variation between listeners or within listeners across different situations.

The following hypotheses were pre-registered, reflecting predictions about each group on its own terms, as well as differences between the groups. For the monodialectals (Hypothesis 1), this constitutes a replication of the findings in Chapter 2, but with a university sample rather than a crowdsourced sample.

The exact text of the pre-registered hypotheses is provided below:

1. Monodialectals will show partial knowledge of AAL morphosyntax.

    A) They will be more likely to perceive regularized subject-verb agreement when an acoustically ambiguous sentence is uttered by an AAL speaker than by an MAE speaker (as measured by transcription of the regularized variant)

    B) They will rate sentences that are grammatical in AAL but not MAE as more likely in AAL than MAE

    C) They will rate sentences that are ungrammatical in both AAL and MAE as more likely in AAL than MAE

2. Bidialectals will show strong knowledge of both MAE and AAL morphosyntax.

    A) They will be more likely to perceive regularized subject-verb agreement when an acoustically ambiguous sentence is uttered by an AAL speaker than by an MAE speaker (as measured by transcription of the regularized variant)

    B) They will rate sentences that are grammatical in AAL but not MAE as more likely in AAL than MAE

    C) They will rate sentences that are ungrammatical in both AAL and MAE as equally unlikely in both dialects

3. Bidialectals will show more strongly differentiated morphosyntactic expectations based on

a speaker's dialect than monodialectals.

A) Bidialectals will be more likely than monodialectals to perceive regularized subject-verb agreement overall (as measured by transcription of the regularized variant)

B) Bidialectals will show a greater differentiation than monodialectals between MAE and AAL in perception. That is, bidialectals will have a steeper slope between the MAE and AAL guises (with more regularized transcriptions for the AAL guise) than will monodialectals.

C) Bidialectals' AAL guise ratings of sentences that are ungrammatical in both AAL and MAE will be lower than monodialectals' ratings.

D) In comparison to monodialectals, bidialectals will show a greater differentiation between guises for regularized subject-verb agreement vs. patterns that are ungrammatical in both MAE and AAL. (In statistical terms, this prediction constitutes a three-way interaction.)

These pre-registered hypotheses reflect a variety of ways that listeners might anticipate the SVA patterns of AAL. However, some alternative hypotheses are plausible. First, bidialectal participants might show the same pattern as monodialectal participants. This could reflect the combination of their linguistic knowledge of MAE and experience with the dominant status of MAE. This would align with work by Sumner and Samuel (2009) on phonetic variation, where there was a general preference for r-ful productions in a priming paradigm, even for participants from NYC who used the r-less variant. Second, bidialectal participants might be generally more linguistically flexible, expecting a variety of speakers to use a variety of forms. This could be observed as less "granular" expectations in the rating task. This would align with work by Bice and Kroll (2019), who found that broad exposure to linguistic diversity makes listeners more flexible and promotes language learning.

3.1.1 Exploratory questions

While all of these hypotheses have been framed in terms of a dichotomous comparison between bidialectal and monodialectal speakers, this likely reflects a simplification. There is widely documented variability in the degree of usage of different features of AAL, both between and within individuals and groups, spanning essentially all variationist work on AAL (e.g., Van Hofwegen and Wolfram, 2017). However, the story is a bit more complex with variation in SVA. Verbal *-s* has been argued to be categorically absent in AAL (e.g., Newkirk-Turner and Green, 2016), though this is disputed (e.g., Baugh, 1990), and the status of auxiliaries is less clear (Maher et al., 2021). Moreover, grammatical AAL features may be less prevalent among middle class speakers and less prone to shifting (Weldon, 2022). Given the mixed evidence on whether a categorical or continuous approach best describes the pattern, it is worth considering whether a more continuous measure of AAL usage/exposure would be more appropriate than the dichotomous approach outlined above. To explore this, I will use participants' ordinal ratings of AAL usage in secondary analyses.

The present study also allows for exploration of the relationship between these types of knowledge. In Chapter 2, the two tasks were administered separately, with different participants sampled from the same population. In this chapter, all participants completed both tasks. This means that we can ask about the relationship between the granularity of participants' knowledge in the rating task and their propensity to predict regularized forms in the transcription task. If these two tasks are both measuring participants' knowledge of SVA variation, we would expect task performance to be correlated at the level of individual participants. Moreover, if prediction in neutral contexts constitutes a higher level of knowledge than recognition, we might expect an implicational hierarchy where only participants who show granular knowledge of AAL grammatical

rules will predict regularized SVA in a neutral context. Conversely, we might find general group effects in each task but no individual-level correlations between tasks. This would prompt careful consideration of how the task demands variably draw upon linguistic knowledge.

3.2 Methods: Experiment 3

Experiment 3 was implemented in Qualtrics and lasted approximately one hour. Two groups of participants—monodialectal speakers of AAL and bidialectal speakers of AAL and MAE—completed the experiment asynchronously on their own devices. After completing a digital consent form, they completed a sentence transcription task (from Experiment 2C), followed by the sentence rating task (Experiment 1). This order was kept fixed across participants so that the explicitly metalinguistic nature of the grammaticality judgment task would not affect the results of the transcription task. After completing both tasks, participants answered open-ended questions about what they noticed about the study. After this, they provided ratings for their usage of and exposure to MAE and AAL. As with the sentence rating task, the two varieties were labeled as "Language Variety 1" (AAL) and "Language Variety 2" (MAE), defined inductively through use of examples with stereotyped features. Participants heard one example from each speaker from the transcription task to remind them of the variety. For Language Variety 1 (AAL), one sentence featured *was*-leveling and one sentence featured negative inversion with negative concord; for Language Variety 2 (MAE), one sentence featured *were* with a plural subject, and another featured single negation without inversion.

Using a four-point scale, participants answered the questions "How often do you interact with people who talk this way?" (exposure) and "How often do *you* talk this way?" (usage). The ends of the scale were labeled as "never" and "all the time," with a separate option for "unsure." An even number of points was chosen to force participants to identify whether they generally do

or do not use a given variety and to avoid the issue where the middle point on a scale serves as a proxy for an "unsure" response (as was found in Experiments 2A and 2B). I selected a small number of points to allow for some variability while acknowledging that fine-grained numerical judgments might be inappropriate for the imprecise nature of defining dialect use, especially when targeting two groups who likely have rather different experiences. Finally, participants completed the demographic and linguistic background questionnaire that was used in Experiments 1 and 2.

3.2.1 Participants

Two groups of participants were recruited: monodialectal speakers of MAE and bidialectal speakers of MAE and AAL. Identification of these groups occurred in two pre-registered stages. First, pre-screening criteria were applied in the targeted advertisement of the study to likely monodialectal and bidialectal speakers. Second, participants answered questions about their linguistic background at the end of the experiment after they had heard examples of both AAL and MAE and were excluded if their responses did not meet the requirements for a given group. The target sample size was 72 participants (36 per group) after pre-registered exclusions. This target sample size is based on a power analysis to ensure 80% power. Because I planned to use mixed-effects regression where power cannot easily be determined analytically, I performed 1000 power simulations based on variance components and parameter estimates from 24-participant pilot experiments, described in Chapters 2. For the sentence rating task, I calculated power as the proportion of simulated datasets where the three-way interaction (among Participant Group, Guise, and Grammaticality Condition) was significant at the $\alpha = 0.05$ level. With a moderate effect corresponding to 1/2 of a Likert-type point, I found approximately 80% power with 72 participants. Similar simulations for the two-way interaction associated with the transcription task indicate approximately 90% power for 72 participants.

3.2.1.1 Recruitment

Regardless of group, participants were only eligible for the study if they grew up in the United States and spoke English at home. For the monodialectal group, participants were eligible if they identified as White and from a middle-class or upper-middle class background, if the version of English they spoke at home was similar to the version of English that is generally used in school, and if their primary caregivers grew up in U.S. and spoke English as their first language. For the bidialectal group, participants self-identified as Black/African American, and the advertising material mentioned "switching between different types of English," as well as the term "African American Vernacular English," given the recent increase in public awareness of this term.

The pre-registered participant pool for the monodialectal group was the UMD paid SONA Psychology pool, where participants would be paid $15 to complete the one-hour task. However, I found that a high proportion of accounts appeared to be fraudulent in this participant pool, with multiple submissions per IP address and several IP addresses outside of the United States. Because of this, data from this participant group were discarded, and I recruited instead from the UMD Psychology credit SONA pool, since each account is assigned to an individual student at the University of Maryland. Each participant received one credit for one hour of participation. The pre-screening criteria were listed in the experiment description on the SONA website, but after some participants who did not meet the racial and linguistic criteria completed the task, pre-screening filters were applied in SONA so that the study would only be displayed to those in the pool who indicated that they were White and spoke English as their first language.

The bidialectal group used a snowball sample, where participants were encouraged to share the sign-up information with friends and family. Participants in this group were paid $25 for the task, since this is a special population that can be more difficult to recruit. A flier was circulated

via email among students at Howard University and Bowie State University, as well as through a program for students of color at the University of Maryland. Additionally, fliers were posted on bulletin boards around the campus of the University of Maryland. Early on in this process, I found that many interested participants had at least one caregiver who did not grow up in the U.S. Because of this, I removed the requirements that participants grew up in a monolingual household and that all caregivers grew up in the U.S. as long as the participant considered English to be one of their first languages. Additionally, I recruited from the UMD credit SONA pool, using pre-screening filters for students who indicated that they were Black and spoke English as their first language.[2]

3.2.2 Filtering participants for analysis

A total of 136 participants completed the study: 66 from the monodialectal recruitment stream and 70 from the bidialectal recruitment stream. Based on pre-registered general exclusion criteria, 19 participants were excluded (non-native speaker of English, incorrect answers to instruction comprehension questions, more than 10% of questions left blank on either task, more than one incorrect response to the guise-establishing audio in the rating task, median string distance from target response greater than 5 for filler items in the transcription task). One participant was excluded due to being older than the maximum age of 50.

Under the pre-registered grouping criteria, participants were to be excluded if they did not speak English as a first language or had another language spoken in their household growing up. Additionally, self-reported usage of and exposure to MAE and AAL was defined according to

[2]Surprisingly, I received eight submissions that appeared to be fraudulent. They were submitted from IP addresses associated with proxy/VPN servers, appeared in sequence, used the international "Date/Month" format in signing the digital consent form, and were not connected to University of Maryland email addresses. I confirmed with the SONA account administrator that these accounts were not associated with any UMD student, and I take them to represent one individual from outside the U.S. who did not understand that the SONA "credit" has no monetary value. These submissions are not included in participant counts.

		MAE use	AAL use	MAE exposure	AAL exposure
Monodialectal	Preregistered	4	1	3 or 4	1 or 2
	Modified	3 or 4	1 or 2	N/A	N/A
Bidialectal	Preregistered	> 1	> 1	> 1	> 1
	Modified	> 1	3 or 4	N/A	N/A

Table 3.1: Criteria for inclusion in monodialectal and bidialectal groups by self-reported usage and exposure ratings (1-4).

the criteria in Table 3.1. Based on these criteria, 13 participants were included in monodialectal group and 33 in the bidialectal group. As we can see in Figure 3.1, despite the low yield rate of these criteria, there are clear differences in the linguistic experience of the two recruited groups. The modal response for the monodialectal recruitment group was 1 for AAL usage and 4 for MAE usage, while the modal response for the bidialectal group was 3 for both AAL and MAE usage. Both groups reported high exposure to MAE (mode = 4). Bidialectal participants reported high exposure to AAL (mode = 4), while monodialectal participants reported middling exposure to AAL (mode = 2, but many responded with 3).

Thus, the pre-registered groupings have an unnecessarily low yield rate that masks a true difference between the groups. Because of this, I created modified groups based primarily on self-reported *use* of each variety, summarized in Table 3.1. These groups were less strictly defined. Exposure criteria were removed, and participants were counted as monodialectal if they were in the lower half of the AAL usage scale and upper half of the MAE usage scale. Since this grouping allowed for monodialectals to report 2/4 on the AAL usage scale, for bidialectals, the range was restricted to 3 or 4 to prevent the ratings from overlapping. Based on these criteria, 36 participants were included in monodialectal group, and 27 participants were included in the bidialectal group.

While reported exposure to each dialect is no longer part of grouping criteria, the relationship between self-reported usage and exposure is summarized in Figure 3.2. In both groups,

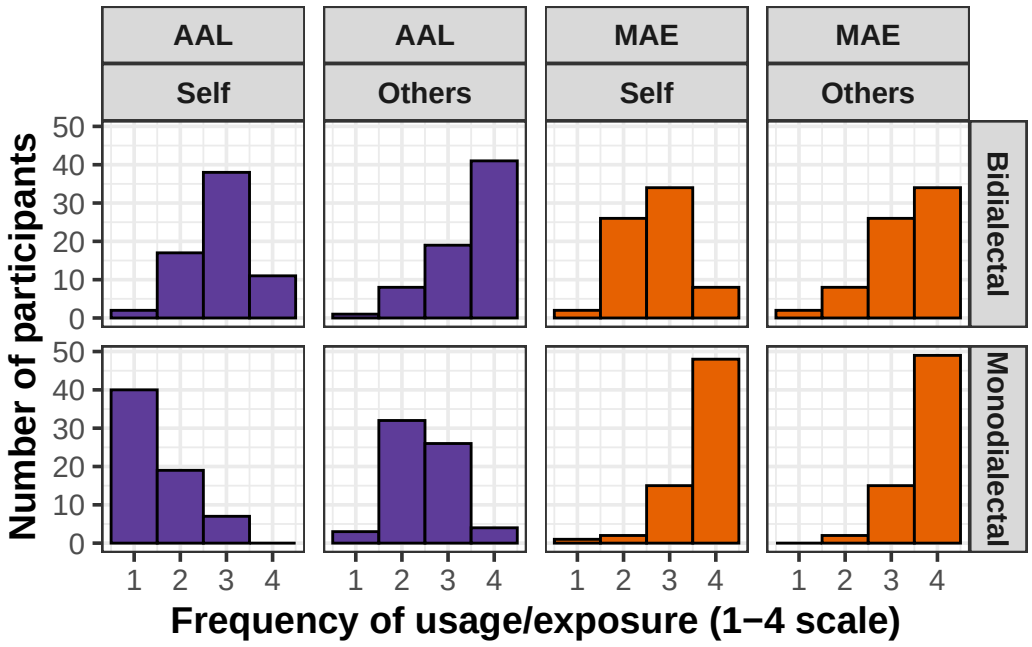

Figure 3.1: Histogram of participant ratings of usage of and exposure to both AAL and MAE on a 1-4 scale. This plot includes all participants, prior to exclusionary criteria.

participants tended to report slightly more exposure than usage to AAL. The modal monodialectal participant reported usage and exposure to MAE at ceiling, and the modal bidialectal participant reported usage level 3 and exposure level 4 for MAE.

For participants who were included in analysis, most participants were female ($n = 20$ for the bidialectal group, $n = 24$ for the monodialectal group), with the remaining participants identifying as male, non-binary, multiple gender identities, or choosing not to disclose. This reflects the skew of our participant pool, as well as the social networks through which the advertisement for the study was distributed to potential bidialectal participants. Monodialectal participants primarily grew up in Maryland ($n = 25$) and bidialectal participants predominantly grew up in Maryland ($n = 10$) or Georgia ($n = 10$). This sample had a narrower age range than the experiments in Chapter 2 due to the university-based sampling procedure (bidialectal: median=20, mean = 22.5, sd = 5.5; monodialectal: median=19.5, mean = 19.7, sd = 1.2).

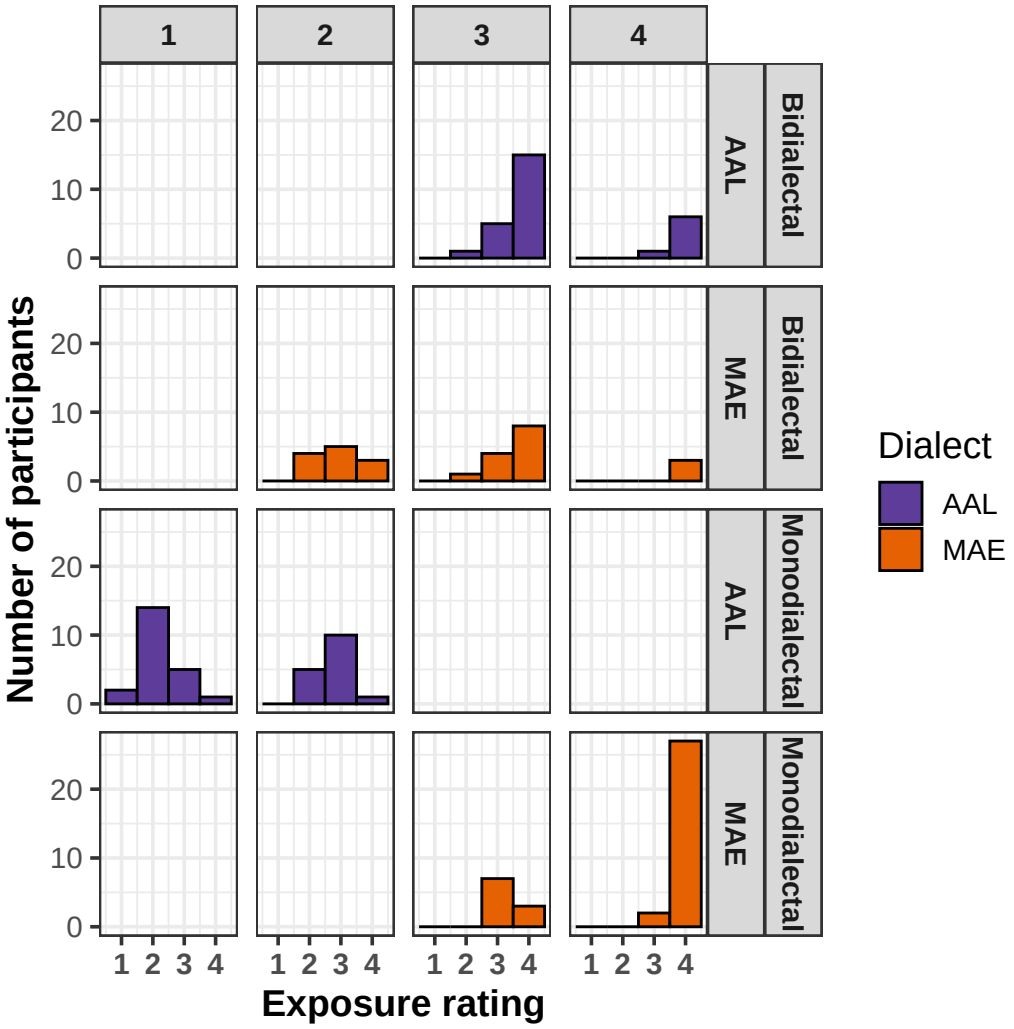

Figure 3.2: Histogram of participants' self-reported exposure to AAL and MAE (x axis), separated by self-reported usage levels (facets). Only participants who met the modified criteria for each group (bidialectal and monodialectal) are included.

3.2.3 Task procedure

3.2.3.1 Rating task

All participants completed a version of the sentence rating task described in Chapter 2, Experiment 1. In this task, participants listen to naturalistic audio examples to identify a given language variety, then rate written sentences for how likely a speaker of that variety would be to say them. Since this task followed the transcription task, I implicitly linked the guises by using a voice from the first part of the experiment as the first example of each guise in the rating task. Otherwise, the same audio from Experiment 1 was used, which was split between targeting morphosyntactic and phonetic features. Participants heard five audio examples at the beginning of each block and answered a multiple choice question identifying one word in the sentence. The target word was where the feature of interest occurred; for example, the target word for (33), the target word was *nothing*. There were two break points per guise block where participants heard two additional examples to remind them of the guise.

(33) I don't like nothing sweet on my chicken.

Items represented three variables associated with AAL: (S), meaning zero-marked simple present tense verbs with third person singular subjects (34), compared to verbal *-s* in MAE; (WAS), referring to the leveling of *was* with plural subjects (35), where *were* would be required in MAE; and non-overt copulas (ZCOP, 36). There were four versions of each item: *Grammatical-Both*, where the morphology is the same in both AAL and MAE; *Grammatical-AAL*, where the morphology is grammatical in AAL, but not in MAE; *Ungrammatical-Agreement*, where the morphology differs from MAE within the agreement (or copula absence) paradigm, but in a way that is ungrammatical in both AAL and MAE; and *Ungrammatical-Other*, where there is a phrase structure error other than the agreement/copula paradigm. A full list of materials can be found

in Appendix B.

(34) **Verbal (S)**

a. **Grammatical-Both:** They like to play a lot.

b. **Grammatical-AAL:** She like to play a lot.

c. **Ungrammatical-Agreement:** *They likes to play a lot.

d. **Ungrammatical-Other:** *Them like to play a lot.

(35) **(WAS)**

a. **Grammatical-Both:** I was crying.

b. **Grammatical-AAL:** They was crying.

c. **Ungrammatical-Agreement:** *I were crying.

d. **Ungrammatical-Other:** *I was cry.

(36) **Null Copula**

a. **Grammatical-Both:** I'm going crazy right now.

b. **Grammatical-AAL:** He going crazy right now.

c. **Marginal-AAL:** The teacher going crazy right now.

d. **Ungrammatical-Agreement:** *I going crazy right now.

e. **Ungrammatical-Other:** *He'll going crazy right now.

All of these sentences are adapted from naturally-occurring examples in CORAAL (Kendall and Farrington, 2021). Each participant saw only one version of each sentence. The final two "Ungrammatical" levels were intended to test the granularity of participants' knowledge of AAL morphology: if participants have highly granular knowledge of the rules of AAL, they will provide high likelihood ratings for Grammatical-Both and Grammatical-AAL items when they are

attributed to AAL, and they will provide low likelihood ratings for both ungrammatical item types. If participants know that SVA rules in AAL are different from those of MAE, they will provide high likelihood ratings for both the Grammatical-AAL and Ungrammatical-Agreement items when these items are attributed to AAL. Finally, if participants simply know that AAL allows for sentences that are "ungrammatical" in MAE, all sentence types—including Ungrammatical-Other—will receive higher likelihood ratings when attributed to AAL than when attributed to MAE. Additionally, as an exploratory measure, I tested whether participants are sensitive to grammatical conditioning effects for the null copula in AAL. In particular, the null copula variant is more common with pronominal subjects (36b) than with full NP subjects (36c), even though it is attested in both environments (Labov, 1969). I will refer to this as the "Marginal-AAL" condition.

For all participants, the AAL block preceded the MAE block to avoid potential order effects. In Experiment 1, this order led to a larger guise effect, presumably because MAE would be the default for most university-based studies, so the contrast is clearer when AAL is presented first. While this is less of a concern when participants have already completed the transcription task, the fixed order facilitates group comparison without an additional between-participants factor. There were ten experimental lists, with each item appearing in at least one list in each grammaticality condition and guise pairing.[3] Items in each list were presented in a fixed, pseudorandom order, and the guise-establishing audio was the same in all lists. Participants were randomly assigned to a list using the Qualtrics randomizer feature, using the option to evenly present lists.

[3]Presentation was slightly unbalanced due to the fact that there were four grammaticality conditions for the verbal (S) and (WAS) items, but five grammaticality conditions for the null copula items.

3.2.3.2 Transcription task

Prior to the rating task, participants completed a version of the transcription task described in Chapter 2, which is adapted from the sociolinguistic repetition task described by Buson et al. (2018). Participants heard four voices: a female AAL speaker, a male AAL speaker, a female MAE speaker, and a male MAE speaker. Since the lect is of primary interest in this study, unless otherwise specified, I will refer to the experimentally-manipulated variety (AAL vs. MAE) as the "guise" manipulation, collapsing across gender. (This differs somewhat from common usage, where the term "guise" would likely apply to each voice-gender pairing.) The same materials and voices were used from Experiment 2C.

As in Experiment 2, critical items had acoustically ambiguous agreement morphology in a context where regularization is common in AAL. This includes sentences prompting (S), meaning that they used simple present tense with a third person singular subject (37a), and sentences prompting (WAS), which involved a past tense copula and plural subject (38a). For (S) sentences, the acoustic ambiguity was the result of ambiguous segmentation, where the word following the verb started with /s/, and the root form of the verb used a voiceless stop, meaning that a potential agreement marker would also be realized with /s/ (instead of the default form, /z/). For (WAS) sentences, most of the *was* token was replaced with white noise, leaving the initial 20 ms to allow for the perception of /w/ at the start of the word; this is the duration that was used in Experiment 2C. All critical items were created using the splicing procedure described in Chapter 2 (e.g., combining 37a and 37b at verb onset) to ensure that the ambiguities were truly equivalent across guises and that the speakers did not record stimuli that violated the grammar of the variety they were representing.

(37) a. He sit(s) still for hours.

b. We sit(s) still for hours.

(38) a. The roads w[NOISE] too slick last night.

b. The road w[NOISE] too slick last night.

Guise-establishing filler items used negative inversion and negative concord for the AAL guise and MAE translation equivalents in the MAE guise. Since the null copula feature was used for the rating task, it was not included here. Additionally, there were agreement filler items where the regular verb form would be used in both varieties: simple present tense with a subject other than third person singular (37b) and past tense copula with a singular subject (38b). These items were included to minimize the risk that participants would learn to always transcribe items with a verbal *-s* or *were*. Half of the filler items contained white noise. For the agreement filler items, *was* was masked by white noise. For the negative inversion and negative concord items, half of the items contained one word with white noise. This was done at varying points in the sentence, replacing most of the vowel of the stressed syllable in a multisyllabic word (e.g., the first vowel of *nobody*), to make the original word easily recoverable and increase the chance that participants form the habit of filling in a word when they hear white noise.

Within a given guise, each participant was presented with 6 critical (WAS) items, 6 critical (S) items, 10 negative inversion items, 8 negative concord items, 3 filler (WAS) items, and 3 filler (S) items. Thus, each participant was presented with 72 total sentences, 36 per guise (and 18 per voice). Critical items were counterbalanced across the four voices in six lists; they were drawn from a pool of 18 (WAS) items and 18 (S) items, since an expanded item pool increases the ability to estimate item-level random effects when a small number of items is presented to a given participant. Lists were randomly assigned using the Qualtrics "randomizer" function, and items were presented in a unique random order for each participant, implemented in Qualtrics. The first eight items were always filler items, with one negative inversion and one negative concord

item per voice, presented in random order; this helped to establish which voices were associated with some non-mainstream grammar before any critical items were presented, as well as providing practice with the task. For each trial, one embedded WAV file appeared on screen, with a space to transcribe the recording. Participants were not prevented from playing the file multiple times, but they could not return to previous trials after progressing to the next page.

3.3 Results

Below, I present the results for the sentence rating and transcription tasks, respectively, and provide exploratory analysis of the relationship between participants performance on the two tasks. All models were fitted according to a preregistered plan using `lme4` (Bates et al., 2015, version 1.1.31) in R (R Development Core Team, 2011), with the `buildmer` package (Voeten, 2022, version 2.7) used determine the optimal random effects structure[4] and `lmerTest` (Kuznetsova et al., 2017, version 3.1.3) used to estimate p values using Satterthwaite's method. *Post hoc* pairwise comparisons were made using the `emmeans` R package (Lenth, 2022, version 1.8.3), using Satterthwaite's method to estimate degrees of freedom and the Tukey method for p value adjustment. Full regression tables and pairwise contrast tables can be found in Appendix B.

3.3.1 Rating task

Results for the sentence rating task are plotted in Figure 3.3 for the modified groupings and in Figure 3.4 for preregistered groupings. The three grammatical phenomena are combined, with the "Marginal-AAL" category excluded.

Several effects are apparent on these plots. There were several similarities between the

[4]In the preregistration, I indicated a plan to use the "backward" fitting approach, where `buildmer` begins with a maximal model and drops terms until the model converges without issues of singular fit. However, I found that the "forward" fitting approach yielded a more complex random effects structure without issues of convergence of singular fit, and since this better accomplishes the goal of building an optimal model, I used this approach.

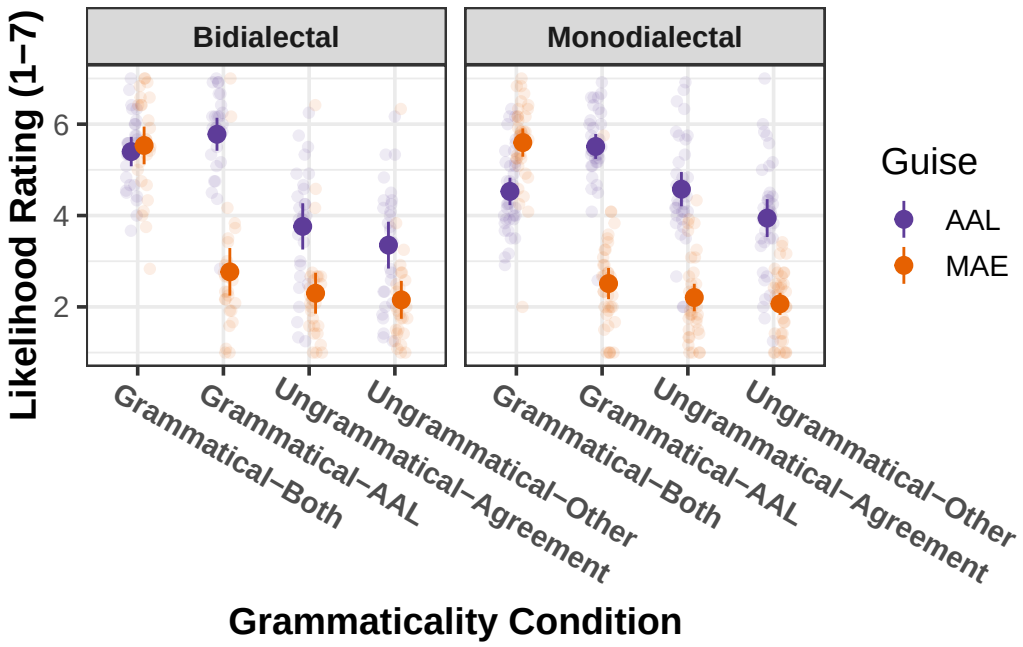

Figure 3.3: Participant ratings in each guise and grammaticality condition for the revised participant groups. Semi-transparent points represent participant-level means, and error bars represent 95% confidence intervals, based on participant-level means.

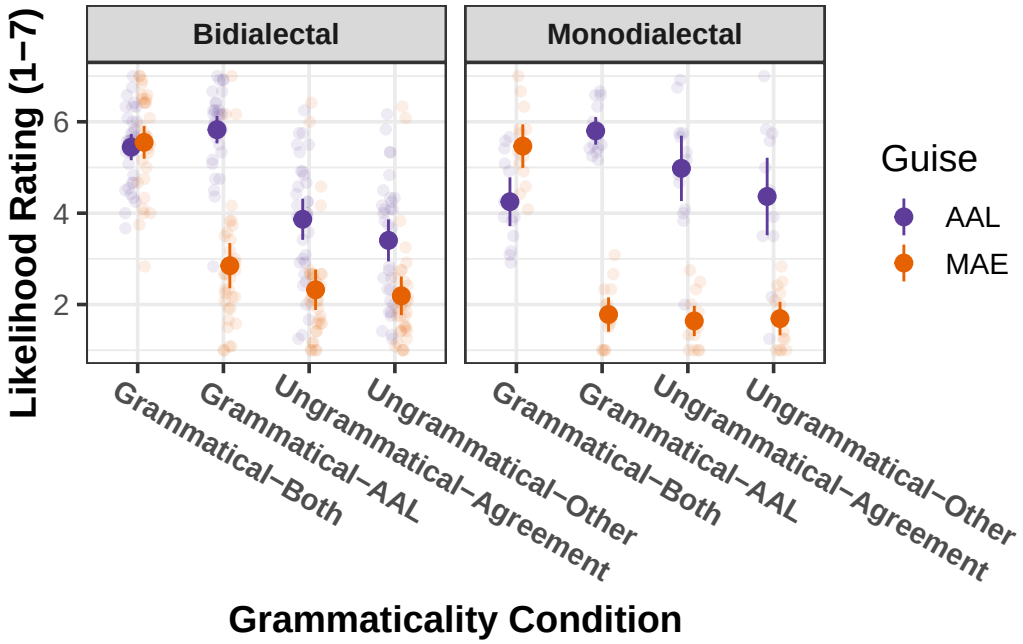

Figure 3.4: Participant ratings in each guise and grammaticality condition for the preregisterd participant groups. Semi-transparent points represent participant-level means, and error bars represent 95% confidence intervals, based on participant-level means.

groups. First, patterns in the MAE Guise reflect expected acceptability judgments: MAE-grammatical sentences are rated as likely for an MAE speaker, and all other grammaticality conditions are rated as unlikely. Second, there is a large difference between guises in the Grammatical-AAL condition, where participants rate sentences as likely for AAL but unlikely for MAE. Third, sentences that are ungrammatical in both AAL and MAE are rated as more likely in AAL than in MAE—even for participants who speak AAL. There are also clear differences between the groups. Monodialectal participants have a stronger expectation of general ungrammaticality in AAL, giving higher likelihood ratings to rate AAL-ungrammatical sentences in the AAL guise than did monodialectal participants. This is particularly clear in the difference between the Grammatical-AAL condition and the Ungrammatical-Agreement condition. Additionally, there was an unexpected difference in the Grammatical-Both condition. Bidialectal participants showed the expected result, where the ratings are the same across guises, while monodialectal participants provided lower likelihood ratings in the AAL guise than in the MAE guise, indicating that they do not expect AAL speakers to utter sentences that are prescriptively grammatical.

These patterns are confirmed by statistical analysis, using pre-registered modeling decisions. To test hypotheses about each participant group on its own terms, I fitted a separate linear mixed-effects model for each group, where guise was dummy-coded, with AAL as the reference condition, and grammaticality condition was forward difference coded, meaning that each level was compared to the previous level, starting from the "Grammatical-Both" condition. For these single-group models, the maximal random effects structure is cross-classified, with both participant- and item-level random slopes, as well as random intercepts for grammaticality condition, guise, and the condition-by-guise interaction for both participants and items. This can be expressed in `lme4` syntax as follow: `(1 + Guise * Condition | ParticipantID) + (1 + Guise * Condition | ItemID)`. For all models, the likelihood rating was recentered at 0 (rather than 4), allowing a

significant intercept term to be interpreted as a non-neutral rating for the reference condition. Regression tables for each model are provided in Appendix B. For ease of interpretation, I will focus on the *post hoc* comparisons of estimated marginal means.

3.3.1.1 Bidialectal group

For the bidialectal group, the optimal random effects structure included participant- and item-level random intercepts and guise-by-participant and guise-by-item random slopes. Pairwise comparisons are provided in Appendix B. For every grammaticality condition, the AAL guise received higher ratings than the MAE guise (ps< 0.001), with the exception of the Grammatical-Both condition, where there was no significant difference ($p = 0.99$). Additionally, comparing adjacent grammaticality conditions within a guise, there is a general trend toward lower ratings moving across the conditions in the order of anticipated grammaticality (Grammatical-Both, Grammatical-AAL, Ungrammatical-Agreement, Ungrammatical-Other; ps< 0.01), with two exceptions: participants gave higher ratings for the Grammatical-AAL condition than the Grammatical-Both condition in the AAL guise ($p = 0.02$), and there was no difference between the Ungrammatical conditions in the MAE guise ($p = 0.88$).

3.3.1.2 Monodialectal group

For the monodialectal group, the optimal random effects structure included participant-level random intercepts, and guise-by-participant and condition-by-participant random slopes. Pairwise comparisons can be found in Appendix B. For every grammaticality condition, the AAL guise received higher ratings than the MAE guise (ps< 0.001), with the exception of the Grammatical-Both condition, where the pattern was reversed; the MAE guise received significantly higher ratings than the AAL guise ($p < 0.001$). Within the AAL guise, participants gave

higher ratings in the Grammatical-AAL condition, relative to the Grammatical-Both condition ($p < 0.001$); lower ratings in the Ungrammatical-Agreement condition ($p < 0.001$), relative to the Grammatical-AAL condition; and lower ratings in the Ungrammatical-Other condition, relative to the Ungrammatical-Agreement condition ($p < 0.001$). Notably, there was no difference between the Grammatical-Both and Ungrammatical-Agreement conditions in the AAL Guise ($p = 1$). In the MAE guise, ratings were significantly lower in the Grammatical-AAL condition than in the Grammatical-Both condition ($p < 0.001$), with no significant difference between any other adjacent grammaticality condition, though there was a significant difference between the Grammatical-AAL and Ungrammatical-Other conditions ($p = 0.04$).

3.3.1.3 Comparing groups

To statistically compare the two groups, I preregistered a simplified model where only the Grammatical-AAL and and Ungrammatical-Agreement grammaticality conditions are included. This tests whether participants have fine-grained knowledge of the subject-verb agreement rules of AAL. Rating scale response (centered at zero) was the dependent variable; independent variables were guise, grammaticality condition, and participant group, as well as all two- and three-way interactions. All independent variables were effects-coded, allowing terms to be interpreted as main effects, with the Grammatical-AAL condition, AAL Guise, and Bidialectal group coded as -0.5, and the Ungrammatical-Agreement condition, MAE Guise, and Monodialectal group coded as 0.5. The optimal random effects structure included participant- and item-level random intercepts and guise-by-participant random slopes.

There was a significant main effect of guise ($\hat{\beta}^{\star} = -0.59$, $p < 0.001$), indicating that participants gave higher ratings in the AAL guise. There was a significant main effect of grammaticality condition ($\hat{\beta}^{\star} = 0.21$, $p < 0.001$), indicating that participants gave higher ratings in

the Grammatical-AAL condition, relative to the Ungrammatical-Agreement condition. There was a significant interaction between guise and condition ($\hat{\beta}^{\star} = -0.12$, $p < 0.001$), indicating a greater difference between the grammaticality conditions in the AAL guise, compared to the MAE guise. There was a significant interaction between group and condition ($\hat{\beta}^{\star} = -0.07$, $p < 0.001$), indicating a greater difference between the grammaticality conditions in the bidialectal group, compared to the monodialectal group. Finally, there was a significant three-interaction of group, guise, and grammaticality condition ($\hat{\beta}^{\star} = 0.05$, $p < 0.001$), indicating that guise-by-condition effect was larger for the bidialectal group than the monodialectal group. There was no significant interaction between guise and group ($\hat{\beta}^{\star} = -0.05$, $p = 0.24$).

3.3.1.4 Comparing grammatical phenomena

The previous analyses collapsed the three grammatical phenomena under study: verbal *-s*, *was*-leveling, and null copula. In this section, I explore potential differences among these phenomena. While (S) and (WAS) are both subject-verb agreement phenomena, there is reason to believe that they pattern differently (Maher et al., 2021), and null copula is a distinct phenomenon that has sometimes received a phonological, rather than fully morphosyntactic, analysis (Labov, 1969). Additionally, recall that unlike the agreement phenomena, null copula has gradient grammatical conditioning effects, where the null variant is more common with pronominal than full NP subjects, but attested with both. Thus, I predicted that bidialectal participants would show these grammatical conditioning effects, rating a null copula as less likely with full a NP subject (condition label: "Marginal-AAL") than with a pronominal subject.

Results are plotted by phenomenon in Figure 3.5. While the general pattern reported for the phenomena together holds across phenomena, it appears that for verbal *-s*, the difference between the Grammatical-AAL and Ungrammatical-Agreement conditions is attenuated for both

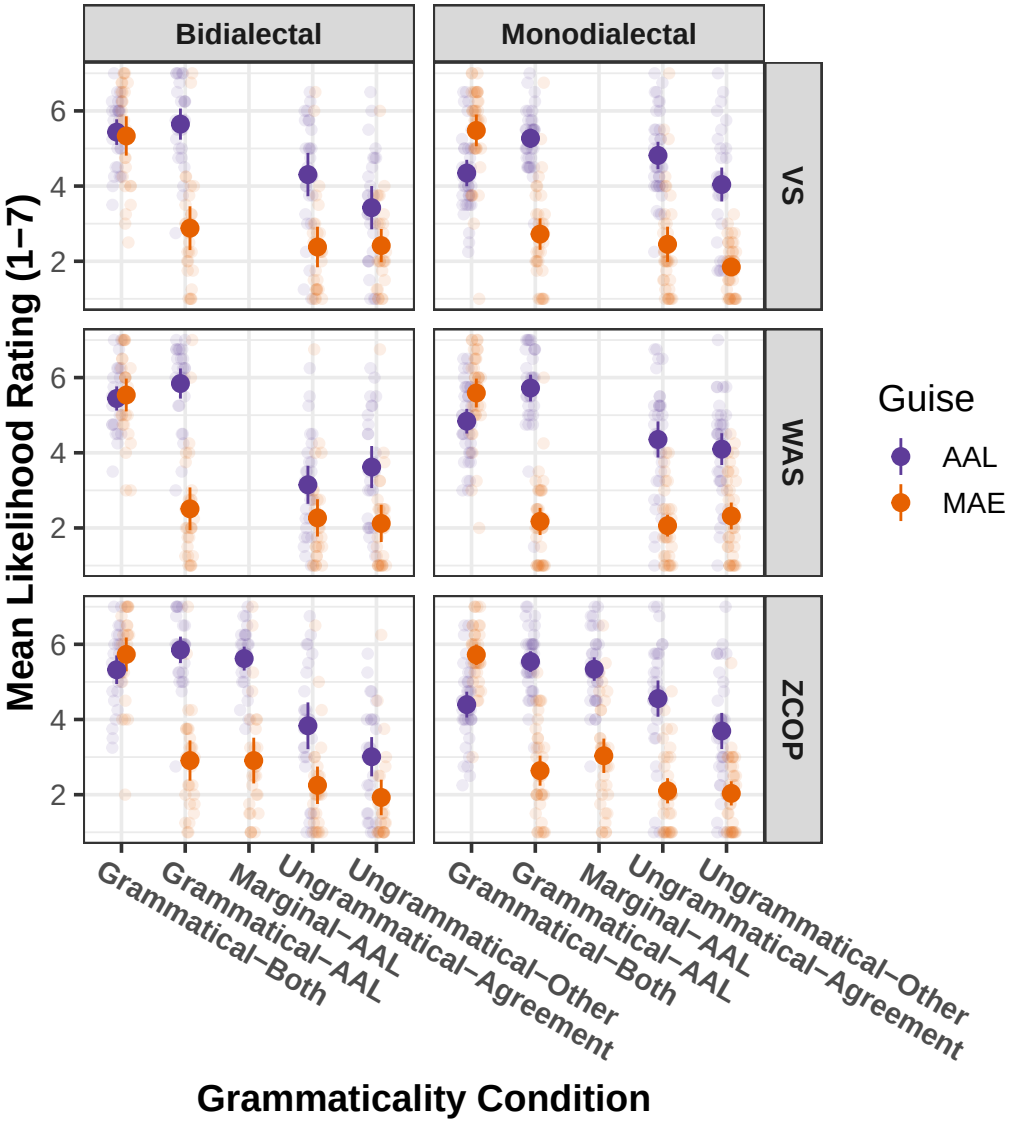

Figure 3.5: Participant ratings by grammmatical phenomenon in each guise and grammaticality condition. VS = Verbal (S), WAS = (WAS), ZCOP = null copula. Semi-transparent points represent participant-level means, and error bars represent 95% confidence intervals, calculated based on participant-level means.

groups, meaning that sentences like *They likes to play a lot* receive higher likelihood ratings than sentences like *I were crying*. For null copula, we see an apparently gradient pattern across grammaticality conditions for the monodialectal group, while for the bidialectal group, the pattern appears categorical, with null copula receiving high likelihood ratings for both pronominal and full NP subjects (excluding *I*), but low likelihood ratings for null copula with *I* as the subject.

Pairwise comparisons for null copula items for each group can be found in the Tables B.9 and B.11. Most importantly for the hypothesis about grammatical conditioning, for both groups, there was no significant difference between the Grammatical-AAL and Marginal-AAL conditions in the AAL guise or in the MAE guise, indicating that participants did not show the grammatical conditioning effects.

Given the visual difference between the two subject-verb agreement phenomena, I performed an exploratory analysis comparing these two phenomena for both participant groups. I first fitted a model with fixed effects of guise, condition, and the guise-by-condition interaction, and the optimal random effects structure, which included participant- and item-level random intercepts, and a guise-by-participant random slope. Then, I tested whether the inclusion of phenomenon in the model (including two-way interactions with guise and condition, as well as a three-way interaction) improved data-model fit. Based on a likelihood ratio test, inclusion of phenomenon in the model significantly improved data-model fit for both the bidialectal group ($X^2(8) = 40.09$, $p < 0.001$) and the monodialectal group ($X^2(8) = 47.58$, $p < 0.001$).

To understand the source of these differences, I performed pairwise comparisons on the full model using `emmeans`. Between-phenomenon comparisons of equivalent guise-grammaticality pairs are provided in Table B.13 for bidialectal participants and in Table B.15 for monodialectal participants (p values reflect adjustments across the 120 possible comparisons). For bidialectal participants, the difference between (S) and (WAS) was only significant in the Ungrammatical-

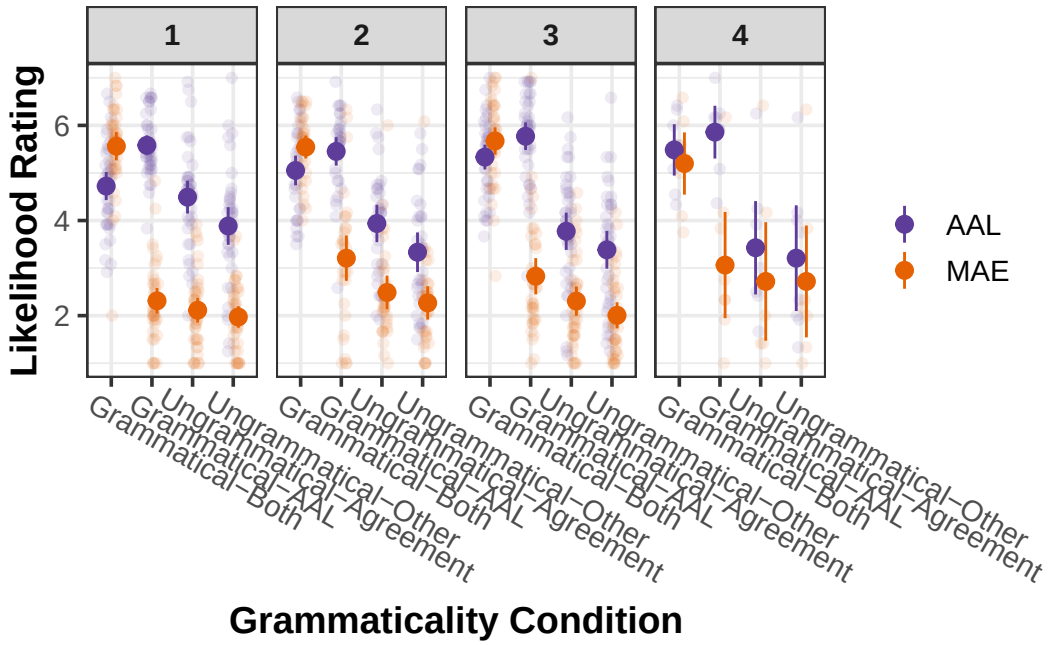

Figure 3.6: Ratings by AAL usage level for all conditions and guises. Facets represent participants' ratings for their own usage of AAL (1=never, 4=all the time). All participants are included, except for those whose first langauge was not English or who were excluded based on data quality filters. Error bars represent 95% confidence intervals, calculated based on participant-level means.

Agreement condition for the AAL guise ($p < 0.001$). This means that for bidialectal participants, sentences like *They speaks Spanish* received higher likelihood ratings than sentences like *He were happy* in the AAL guise. For the monodialectal group, there were no such differences between SVA phenomena.

3.3.1.5 Continuous analysis of group

While the planned analyses involved two groups, participants' self-ratings of AAL usage also allow for a more continuous analysis, without excluding participants whose ratings place them between the categories of "monodialectal" and "bidialectal." In Figure 3.6, participants' ratings are plotted by self-reported AAL usage on a 1-4 scale, and Figure 3.7 shows participant-level means for the contrast of particular interest: the Grammatical-AAL and Ungrammatical-Agreement conditions in the AAL Guise. We can see an increasingly negative slope between

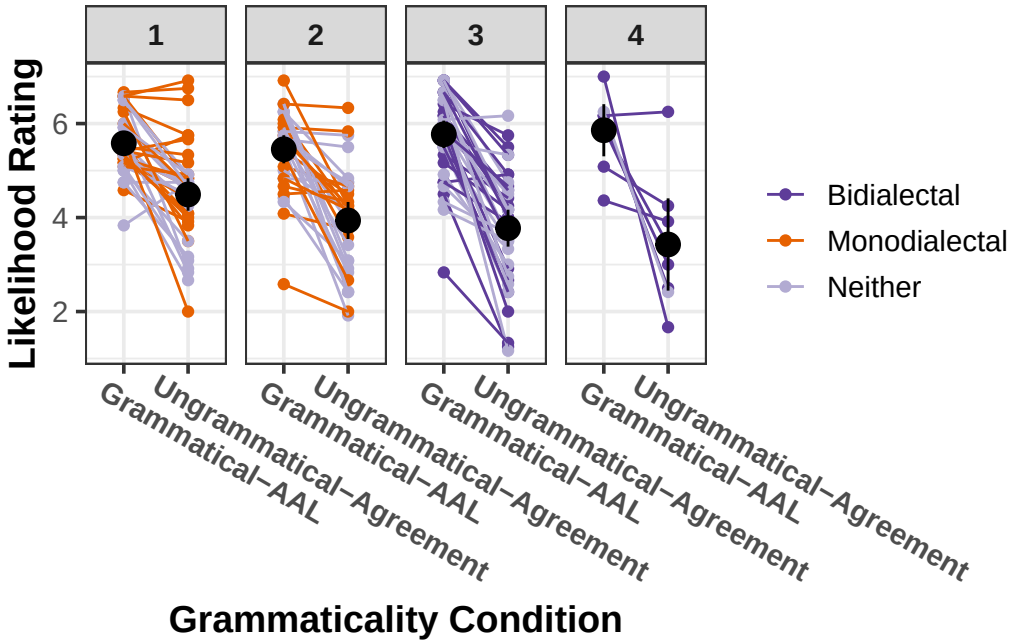

Figure 3.7: Participant-level effects for the two-way condition contrast in the AAL guise. Each line represents a participant, and the black points represent grand means across participants. Facets represent participants' ratings for their own usage of AAL (1=never, 4=all the time). All participants are included, except for those whose first langauge was not English or who were excluded based on data quality filters. Error bars represent 95% confidence intervals, calculated based on participant-level means.

these conditions as participants report greater AAL use, mirroring the effect in the dichotomous comparison between monodialectals and bidialectals.

This effect was tested using a linear mixed-effects model, where rating (centered at 0) was the dependent variable, with fixed effects for AAL-self rating (forward difference coded, with 1 as the reference), condition (effects coded), and the interaction of these effects. The optimal random effects structure included random intercepts for participant and item and a participant-by-grammaticality condition random slope. There was a significant effect of condition ($\hat{\beta}^{\star} = 0.5$, $p < 0.001$), indicating higher ratings in the Grammatical-AAL condition than in the Ungrammatical-Agreement condition for the reference group. For the difference code comparisons, there was a marginal difference in average rating between grammaticality conditions between AAL-self usage level 1 and 2 ($\hat{\beta}^{\star} = 0.19$, $p = 0.08$), but no difference between levels 2 and 3 ($\hat{\beta}^{\star} = -0.04$, $p = 0.7$) and levels 3 and 4 ($\hat{\beta}^{\star} = 0.08$, $p = 0.66$). There was a marginal interaction term, indicating a steeper slope between conditions as self-reported AAL usage increases between usage level 2 and 3 ($\hat{\beta}^{\star} = -0.14$, $p = 0.09$), but not between levels 1 and 2 ($\hat{\beta}^{\star} = -0.11$, $p = 0.17$) or between levels 3 and 4 ($\hat{\beta}^{\star} = -0.13$, $p = 0.34$).

3.3.1.6 Interim discussion: Rating task

As hypothesized, both monodialectal and bidialectal participants rated sentences that are grammatical in AAL but not MAE (Grammatical-AAL condition) as more likely in AAL than in MAE. This demonstrates recognition of the morphosyntactic phenomena, as well as indexical knowledge linking regularized SVA and null copula to AAL. The evidence about the granularity of this knowledge is mixed. For both groups, sentences that are ungrammatical in AAL (Ungrammatical-Agreement and Ungrammatical-Other) received higher ratings in the AAL guise than in the MAE guise, but bidialectal participants showed greater differentiation in the AAL

between the Grammatical-AAL and Ungrammatical-Agreement conditions, compared to monodialectals. This indicates a combination of knowledge that particular grammatical patterns are common in AAL and a general expectation of ungrammaticality in AAL.

Bidialectal participants' expectations of ungrammaticality in AAL. It is a bit surprising that participants who speak AAL would nevertheless attribute ill-formed sentences to AAL. There are several reasons why this might be the case. First, standard language ideology (Lippi-Green, 1997) is common even among AAL speakers, meaning that they might view AAL as "incorrect" even though they speak it. Thus, in this highly metalinguistic task, their responses might reflect a combination of their knowledge of AAL's actual grammar and a belief that AAL generally "doesn't follow grammatical rules." Second, AAL speakers might appreciate the diversity in AAL and want to allow for the possibility that other speakers will have different rules. There was one comment to this effect, where a bidialectal participant suggested that older southern speakers might use *I* with a null copula, even though the participant would not do so themselves.[5] The audio clips deliberately illustrated a range of AAL speakers from several regions and age cohorts, so it is possible that at least some bidialectal participants thought of the AAL guise as a broader category than the version of AAL spoken in their communities. It is still noteworthy, though, that they did not make an equivalent generalization in the MAE guise, as this also included a range of speakers. Third, it is possible that the fixed order of the guises played a role. Since the AAL guise always preceded the MAE guise, it is possible that participants had a better sense of the range of acceptability that could be found in the task when they reached the MAE guise. Thus, by the time they reached the MAE guise, they were more likely to use the end of the scale. This seems unlikely, though, since the same qualitative pattern was observed for either guise order in Experiment 1.

[5]I am not aware of any research suggesting that this is true, although there are many creoles with non-overt copulas, including among immigrant communities.

Finally, perhaps some of the "ungrammatical" sentences actually were within the grammar of some speakers of AAL. This is particularly plausible for verbal *-s* with plural subjects, which received higher ratings than other Ungrammatical-Agreement sentences. Some researchers, such as Baugh (1990), have questioned the assessment that verbal *-s* lies outside of the AAL grammar, and even those who argue that verbal *-s* is categorically absent reference the fact that it is sometimes attested in broader contexts, arguing that this is a hypercorrection in an MAE-dominated society (e.g., Labov and Baker, 2015). This means that AAL speakers have probably been exposed to sentences like *They likes to play* and therefore give them higher ratings. While this played a role in the broader result, it does not fully explain the pattern of ungrammatical sentences receiving higher ratings in the AAL guise, as this pattern was also observed for other items, including items with larger phrase structure errors in the Ungrammatical-Other condition.

Grammatical conditioning for the null copula. The effect for verbal *-s* in the Ungrammatical-Agreement condition might also bear on the research questions about sensitivity to grammatical conditioning. Under an analysis where verbal *-s* is an optional tense marker, the effect observed in the bidialectal speakers reflects sensitivity to the grammatical conditioning of this variable. For the null copula, I hypothesized that there would be a grammatical conditioning effect by subject type with pronominal subjects (the Grammatical-AAL condition) receiving higher likelihood ratings than full NP subjects (the Marginal-AAL condition). Instead, there was no difference between these two conditions for either participant group; bidialectal participants showed an essentially categorical effect, with high likelihood ratings in a permissible environment and low likelihood ratings in other environments, while monodialectal participants showed a more gradient effect, but without a significant difference between the Marginal-AAL condition and adjacent grammaticality conditions. This contrasts with the work of Bender (2005), who found that participants with some knowledge of AAL make stronger social judgments for a null copula

in an uncommon environment, compared to a common environment. However, there are several meaningful differences between her design and mine. She contrasted *predicate* types (VP vs. NP, where VP is more common), while I tested *subject* types. Additionally, her task was a sociolinguistic perception task, where participants rated sentences on various social attributes, given a linguistic feature; I was instead asking participants to rate on linguistic expectations, given other associated sociolinguistic features. Moreover, in both studies, the findings were exploratory, so more work is necessary to understand participants' knowledge of the grammatical conditioning of variable morphosyntactic phenomena.

As with any null effect, multiple interpretations are possible. If this is a true null effect, it could reflect a difference between listeners' expectations and the true usage patterns, meaning that these participants' exposure to AAL aligns with the documented pattern, but given the level of variability, they expect any option that is consistent with AAL grammar. It could also reflect a difference between the documented pattern and participants' experience with AAL; perhaps the reported difference between pronominal and full NP predicates does not reflect these participants' AAL input.[6]

Alternatively, it is plausible that these participants do adjust their expectations regarding copula absence depending on the grammatical environment, but this task was not sufficiently sensitive. Participants were rating these sentences on a Likert-type scale without directly comparing them, and the "worst" items contained phrase structure errors. Moreover, participants were asked how likely the presented form would be, rather than saying whether they would prefer an alternative form. (While the availability of a preferred alternative form might play a role in this likelihood rating, it was probably indirect and also at play with the other forms that differ from MAE. Just as *The teacher going crazy* might be implicitly compared to *The teacher is going*

[6]This could be tested by using the recent extensive data for Washington, D.C. in CORAAL, but that is beyond the scope of this dissertation.

crazy, *She going crazy* might be implicitly compared to *She is going crazy*). Thus, while showing clear contrasts for less subtle effects, the task might have muted any difference that could have otherwise been observed for the grammatical conditioning of the copula. Future work could use different designs to address this issue, such as the binary comparison approach for habitual *be* used by Wolfram (1982), where participants directly compare two sentences and select the one where a given form would best fit, or the rating approach used by Vaughn and Kendall (2018) for (ING), where participants rate the relative likelihood of each variant on a Likert-type scale.

Effects for items that are grammatical in both MAE and AAL. I had not made any predictions about differences about the Grammatical-Both condition, beyond the general expectation that it would receive high likelihood ratings for both guises. However, for both participant groups, participants give *lower* ratings for the AAL-Guise in the Grammatical-Both condition than in the Grammatical-AAL condition; that is, they rated *I was crying* as more likely from an AAL speaker than *They was crying.* For monodialectal participants, this difference was of similar magnitude to the difference between the Grammatical-AAL and Ungrammatical-Agreement conditions, meaning that they thought AAL speakers were equally likely to say *I was crying* and *He were crying.* Additionally, Grammatical-Both items received higher likelihood ratings in the MAE guise than in the AAL guise, but only for monodialectal participants.

On one level, this pattern is surprising, as it clearly reflects a difference between participants' grammatical expectations and the grammatical patterns in the world. However, it brings into focus two different understandings of what AAL is. In one conceptualization, AAL is a full rule-governed grammatical system, meaning that anything licensed by its grammar is AAL, regardless of whether it overlaps with MAE (e.g., Green, 2002). This roughly corresponds to a "first wave" framing of sociolinguistics, where the goal is to characterize the linguistic systems of different social groups, defined based on broad demographic factors (albeit typically with a focus on differences

from dominant forms). An alternative framing would be that AAL is defined by the linguistic features that specifically index a Black/African American identity. In this view *They was* carries a social meaning associated with being African American, while *He was* does not. This latter approach more closely aligns with a "third wave" approach to sociolinguistics, which places the focus on contextually-defined indexical meanings (Eckert, 2012).

The pattern observed here shows that participants' judgments were a mix of both linguistic judgments (what is grammatical for a given speaker) and sociolinguistic judgments (what variants index the social category assigned to the speaker). It seems that monodialectal speakers have linguistic expectations in situations where there is a clear indexical link—i.e., stereotyped features of AAL—but less clear linguistic expectations otherwise. Bidialectal speakers have a clearer representation of AAL as a linguistic system, in addition to having specific indexical values for regularized forms that differ from MAE. This is reflected in the equally high likelihood ratings for Grammatical-Both items in the AAL and MAE guises, as well as the slightly higher ratings for Grammatical-AAL items, reflecting a boost from the indexical value of these items.

3.3.2 Transcription task

Transcriptions were coded for whether the participant used regularized SVA (e.g., *She speak Spanish*), MAE SVA (e.g., *She speaks Spanish*), or did not fit either expected response (e.g., *She spoke Spanish*). These codes were generated automatically using a string match for the subject-verb sequence (e.g., "She speak(s)"), ignoring capitalization and punctuation. Proportions of missing data are provided in Table 3.2. There were more unscorable items in the AAL guise than in the MAE guise, more unscorable items for (WAS) than (S), and more unscorable items for the monodialectal group than the bidialectal group. This makes sense, since (WAS) items involved white noise replacing most of a word, meaning that participants might fill in the blank with

another word or simply omit the word. It also makes sense that participants would have lower transcription accuracy in the AAL guise, considering that it is less familiar to many of them.

Table 3.2: Proportion of unscorable items in the sentence transcription task

Group	Phenomenon	AAL	MAE
Bidialectal	verbal S	0.07	0.02
Bidialectal	was/were	0.17	0.14
Monodialectal	verbal S	0.08	0.05
Monodialectal	was/were	0.13	0.11

To test whether these differences in missingness across conditions was statistically significant, I fitted a logistic mixed-effects model. The dependent variable was whether or not the item was missing (1 = not scorable, 0 = scored as regularized or MAE SVA), and there were fixed effects of group, guise, and SVA phenomenon, all of which were effects coded (-0.5 for the bidialectal group, AAL guise, and verbal *-s*; 0.5 for the monodialectal group, MAE guise, and (WAS)), as well as all possible two- and three-way interactions. The optimal random effects structure included random intercepts for participant and item, but no random slopes. There was a significant effect of guise ($\hat{\beta}^{\star} = -0.28$, $p = 0.007$), meaning that items were more likely to be unscorable in the AAL guise, compared to the MAE guise, and there was a significant effect of SVA phenomenon ($\hat{\beta}^{\star} = 0.53$, $p < 0.001$), meaning that (WAS) items were more likely to be unscorable than (S) items. There was marginal group-by-phenomenon interaction ($\hat{\beta}^{\star} = -0.18$, $p = 0.09$), potentially indicating that the difference between the (S) and (WAS) items was larger for bidialectal participants than for monodialectal participants. The following terms were not significant: participant group ($\hat{\beta}^{\star} = 0.05$, $p = 0.72$), group-by-guise interaction ($\hat{\beta}^{\star} = 0.06$, $p = 0.55$), guise-by-phenomenon interaction ($\hat{\beta}^{\star} = 0.14$, $p = 0.17$), and three-way interaction ($\hat{\beta}^{\star} = -0.07$, $p = 0.54$).

Results from the transcription task are plotted in Figure 3.8, which summarizes the data

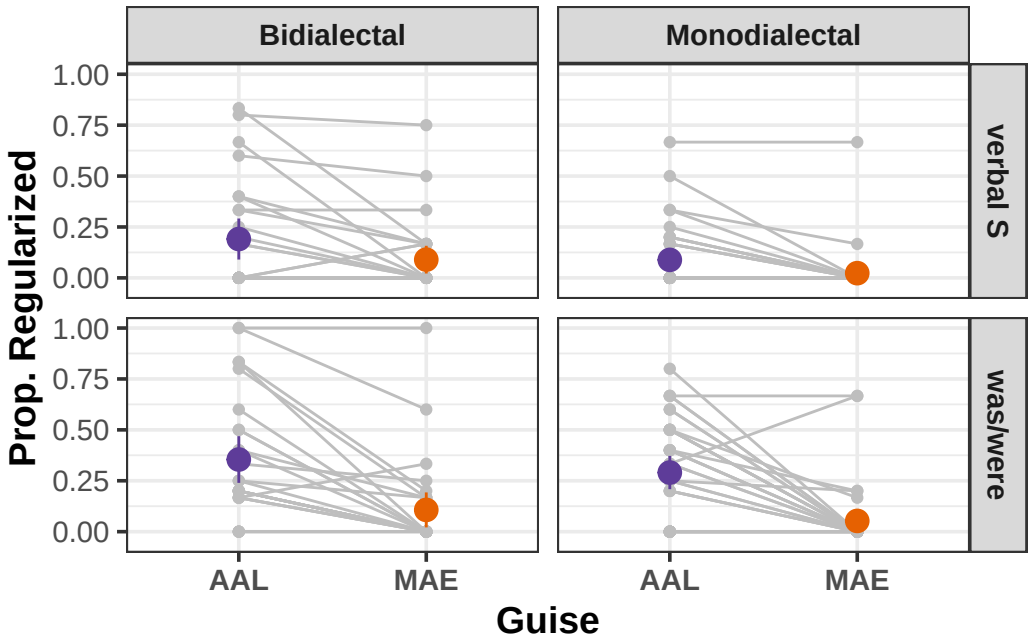

Figure 3.8: Regularized responses in the transcription task, as a proportion of scorable items. Gray lines and points represent participant-level proportions, with grand means provided in color. Error bars represent 95% confidence intervals, calculated based on participant-level means. Note that a given line might represent more than one participant, particularly in the case of participants who never made a regularized transcription, regardless of guise.

by group, guise, and SVA phenomenon. There was an apparent difference between guises, where participants were more likely to transcribe regularized SVA in the AAL guise, compared to the MAE guise, and an apparent difference between groups, where bidialectal participants transcribed with regularized SVA more frequently than monodialectal participants.

Table 3.3: Proportion of participants with each type of guise effect in each group.

Response Type	Bidialectal	Monodialectal
Higher AAL	0.81	0.72
Higher MAE	0.07	0.03
No regularized responses	0.11	0.25
No guise effect (other)	0.00	0.00

These differences were evaluated statistically using logistic mixed-effects models, where the dependent variable was whether a given transcription used the regularized variation, compared to

using the MAE variant.[7] The optimal random effects structure included participant- and item-level random intercepts and guise-by-participant and guise-by-item random slopes. I used a pre-registered model building process where I began with effects-coded fixed effects of interest: guise, group, and the guise-by-group interaction. I then added SVA phenomenon (including all possible interactions) to the model, which significantly improved data-model fit based on a likelihood ratio test ($X^2(4) = 17.88$, $p = 0.001$). Guise gender (male vs. female) did not further improve data-model fit ($X^2(8) = 8.51$, $p = 0.39$). Thus, I report the results for the model including guise, participant group, SVA phenomenon, and all possible two- and three-way interactions.

There was a significant intercept ($\hat{\beta} = -3.07$, $p < 0.001$), meaning that participants transcribed regularized SVA less than half of the time. There was a significant effect of guise ($\hat{\beta} = -2.69$, $p < 0.001$), meaning that participants were less likely to transcribe regularized SVA in the MAE guise, compared to the AAL guise. There was a significant effect of group ($\hat{\beta} = -0.99$, $p = 0.05$), meaning that monodialectal participants were less likely to transcribe regularized SVA, compared to bidialectal participants. Finally, there was a significant effect of SVA phenomenon ($\hat{\beta} = 1.04$, $p = 0.001$), meaning that regularized transcriptions were more likely for (WAS) items than (S) items; in other words, *was* with a plural subject (*They was crying*) was more likely than verbal *-s* absence with a third person singular subject (*She like to play*). No other effects were statistically significant. Contrary to my hypotheses, there was no guise-by-group effect ($\hat{\beta} = -0.68$, $p = 0.32$), meaning that there was no difference between participant groups' level of differentiation between the two guises. There was a marginal interaction between group and SVA phenomenon ($\hat{\beta} = 0.82$, $p = 0.08$). There was no interaction between guise and SVA phenomenon ($\hat{\beta} = -0.79$, $p = 0.17$) and no three-way interaction ($\hat{\beta} = 0.26$, $p = 0.78$).

[7]For the reported results, any other response was treated as missing data. However, identical inferences can be made from a model where missing data is instead included as a non-regularized response, meaning that it is coded as zero for the model, alongside MAE variants.

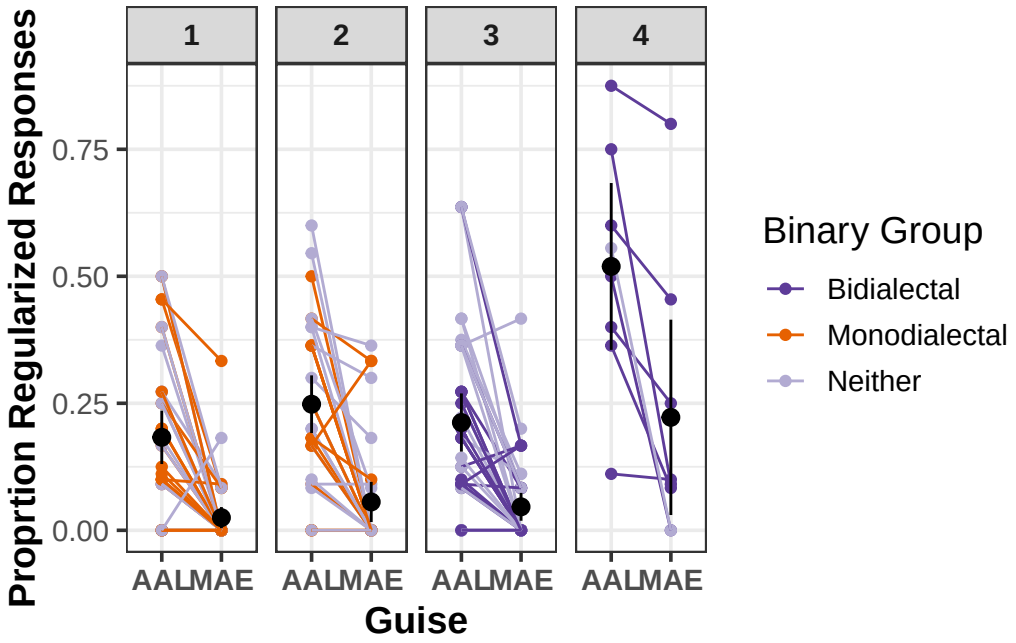

Figure 3.9: Regularized transcriptions by participants' AAL usage

3.3.3 Guise effect by AAL usage level

As with the sentence rating task, I performed exploratory analysis of the guise effect by participants' self-reported usage of AAL on a 4-point scale, allowing for inclusion of participants who did not fit into the binary definitions of "monodialectal" and "bidialectal." These results are plotted in Figure 3.9. We can observe an overall increase in regularized transcriptions as participants increase their self-reported AAL usage.

These results were tested using a logistic mixed-effects model. AAL self report was forward-difference coded with 1 as the reference level, and guise was effects-coded. There was a significant intercept ($\hat{\beta}^{\star} = -2.7$, $p < 0.001$), meaning that participants transcribed regularized SVA less than half of the time. There was a significant effect of guise ($\hat{\beta}^{\star} = -1.42$, $p < 0.001$), meaning that participants were more likely to transcribe regularized SVA in the AAL guise, compared to the MAE guise. There was a marginal difference between AAL usage levels 1 and 2 ($\hat{\beta}^{\star} = -0.78$, $p = 0.06$), meaning that participants with usage level 2 were potentially more likely to

make regularized transcriptions than participants with usage level 1. There was *not* a significant difference between AAL usage levels 2 and 3 ($\hat{\beta}^{\star} = 0.22$, $p = 0.59$). Finally, there was a significant difference between AAL usage levels 3 and 4 ($\hat{\beta}^{\star} = -2.22$, $p < 0.001$), meaning that participants with usage level 4 were more likely to make regularized transcriptions than participants with usage level 3. There were no significant interaction terms, meaning that there was no difference in the magnitude of the guise effect as we move from usage level 1 to 2 ($\hat{\beta}^{\star} = -0.2$, $p = 0.56$), 2 to 3 ($\hat{\beta}^{\star} = -0.09$, $p = 0.76$), or 3 to 4 ($\hat{\beta}^{\star} = -0.2$, $p = 0.65$).

3.3.3.1 Transcription of uncommon SVA

Table 3.4: All transcriptions using subject-verb agreement patterns that are considered ungrammatical in both AAL and MAE.

Group	Guise	Response
Monodialectal	MAE	we breaks some of the things away
Bidialectal	AAL	The girls creates something pretty
Bidialectal	MAE	My parents keeps stuff up there
Bidialectal	AAL	my parents keeps stuff up there

In the rating task, some participants indicated that AAL speakers might use uncommon forms such as verbal *-s* with subjects other than third person singular or *were* with a plural subject. But do participants expect such forms in AAL when they are presented in a neutral context? This experiment included filler versions of the experimental items, where the (WAS) items had a singular subject and the (S) items had a subject other than third-person singular. For these items, only *was* and zero-marking, respectively, are grammatical in either dialect. (There were twice as many critical items as filler items of this type.)

To test whether participants provided transcriptions with these uncommon forms, I used regular expressions to find responses where (WAS) items included the string "were" (surrounded

by spaces) in contexts where it was not preceded by *you*, *we*, *they*, or a word ending in *s*, in addition to responses with (S) items, where the verb was transcribed with an "s" at the end. This yielded only 4 (out of 756) trials, all of which are provided in Table 3.4. There was an even split by guise for this rare response type. This suggests that although participants indicated such items were more likely in AAL in the rating task, they do not expect them when given a neutral option.

3.3.3.2 Interim discussion: Transcription task

As hypothesized, I found an overall effect of guise, where participants in both groups were more likely to transcribe regularized subject-verb agreement in the AAL guise, compared to the MAE guise. In this task, an MAE-grammatical alternative was always consistent with the acoustic signal, so this means that participants were drawing on some knowledge of AAL to resolve the acoustic ambiguity. At the same time, participants maintained a general preference for MAE-grammatical transcriptions, with the majority of responses aligning with MAE grammar, even in the AAL guise. I will return to this point in the general discussion for this chapter.

Contrary to my hypotheses, the guise effect did not statistically differ between groups, as reflected in a non-significant interaction. Instead, there was only a main effect of group, where bidialectal participants were generally more likely to make regularized transcriptions—including in the MAE guise. Taken together, this means that in this task, all participants expected regularized SVA to a greater degree in AAL than MAE, but bidialectal participants viewed regularized SVA as generally more likely, regardless of speaker type. This might reflect that this variant is generally more available to them, particularly in a context where both dialects are in use. However, this finding differs from the three-way interaction effect observed in the rating task, which I consider further below.

Effect of SVA phenomenon. Beyond the effects of primary interest, there was also an effect of item type, where regularized transcriptions were more likely for (WAS) than (S), which replicates findings from Experiments 2A and 2B with participants on Prolific. As noted in Chapter 2, this could reflect a variety of factors. From the perspective of a participant, there was a salient difference between these item types, as (WAS) items used white noise to produce the acoustic ambiguity, while (S) items relied on a segmentation ambiguity and could go entirely unnoticed. This could have affected participants' behavior with the items and led to more explicit reasoning about the absent word in the case of (WAS) items. Additionally, while both item types reflect common SVA patterns, there are many linguistic differences that could affect both their representation and task performance: /s/ is one phone and a bound morpheme, while *was* is three phones and a free morpheme. Both of these factors align with the salience of the white noise in the task to increase attention to *was* and increase the chance that metalinguistic judgments are involved in the transcriptions.

It is also possible that there is a frequency difference between these SVA phenomena, meaning that the observed effect reflects usage in the world. There are multiple senses of "frequency" that might differ. Thinking in variationist terms, we might evaluate frequency of the regularized variant *within the environment where it could appear*. This would imply that *was* with plural subjects is more common than zero marking of simple present tense verbs with third person singular subjects. It is unclear whether this is the case, though Maher et al. (2021) found that child speakers of AAL are more likely to use *were* than overt /s/, which is the opposite pattern. Alternatively, the relevant frequency difference could be how often each regularized form appears *as a proportion of all speech*. As forms of the copula, both *was* and *were* are almost certainly higher frequency than most of the (S) verbs (even though these were also high-frequency verbs), though this raises an additional question of whether the relevant frequency would be for a given

verb or for any verb in simple present tense with a third person singular subject. An answer to this question is beyond the scope of this dissertation, but an interesting area for future research on the grammatical conditioning of SVA variation.

Accuracy/missingness effect. In addition to being more likely to transcribe regularized SVA in the AAL guise, compared to the MAE guise, participants were more likely to provide a transcription that did not align with either expected option, generally indicating a transcription error, or in the case of (WAS), transcribing nothing for the noise-masked word or filling in a word other than *was* or *were*. Interestingly, there was no significant effect of participant group, indicating a general advantage for MAE speech. This replicates previous results, such as those of Jones et al. (2019), who found low accuracy rates when court reporters transcribed AAL, with no impact of participant race or familiarity with AAL on transcription accuracy.[8] This might partially reflect familiarity, where both speaker groups have considerable exposure to MAE due to its dominant status (see Sumner and Samuel, 2009), and it might also reflect ideologies about intelligibility, as listeners' expectations affect transcription accuracy of the same materials (Vaughn, 2019). Regardless, this finding does not affect our interpretation of the effects of primary interest, since changing the denominator for the regularized responses did not affect which terms in the model were significant.

Defining group. While an effect was observed for AAL usage levels, the relevant differences fell *within* the categories used to define monodialectal vs. bidialectal participants rather than *between* the categories; that is, there was no difference between usage levels 2 and 3, but there was a (marginal) difference between levels 1 and 2 (within the monodialectal criteria), as well as between levels 3 and 4 (within the bidialectal criteria). This suggests that the participants at the extremes of each group drove the effect between groups. This also differs from the findings for

[8] I am thankful to Sarah Phillips for pointing out the relevance of this finding to the present results.

the rating task, which I discuss in further detail below.

3.3.4 Combined analysis of the two tasks

If both tasks are tapping into a shared pool of (socio)linguistic knowledge, we might expect performance on the two tasks to correlate. To explore the relationship between participants' performance on the two tasks, I compared participant-level variability in the effects of interest for each task. For the rating task, this is the level of differentiation between the Grammatical-AAL condition and Ungrammatical-Agreement condition in the AAL guise. For the transcription task, this is both the general tendency to transcribe regularized SVA (the main effects found above) and the degree to which this is differentiated by guise (the hypothesized interaction that was not observed above).

I re-fitted each model without the effect of group, since the goal was to characterize individual, rather than group-level variability. This means that the model for the rating task included a fixed effect of grammaticality condition (Grammatical-AAL, Ungrammatical-Agreement), using using only the subset of the data for the AAL guise, and the model for the transcription task included fixed effects of guise, SVA phenomenon, and the two-way interaction, using data from both guises. Both models were cross-classified by participant and item, with random slopes by grammaticality condition for the rating task and random slopes by guise for the transcription task. Correlations between these random effects are plotted in Figure 3.10.

There was a significant correlation between the slopes and intercepts within the rating task ($r = -0.47$, $t(115) = -5.73$, $p < 0.001$) and the transcription task ($r = 0.72$, $t(115) = 11.23$, $p < 0.001$). This reflects the fact that slope-intercept correlations were estimated in the model. More interestingly, there was a significant correlation between participant-level slopes for the rating task and participant-level intercepts for the transcription task ($r = 0.24$, $t(115) = 2.68$,

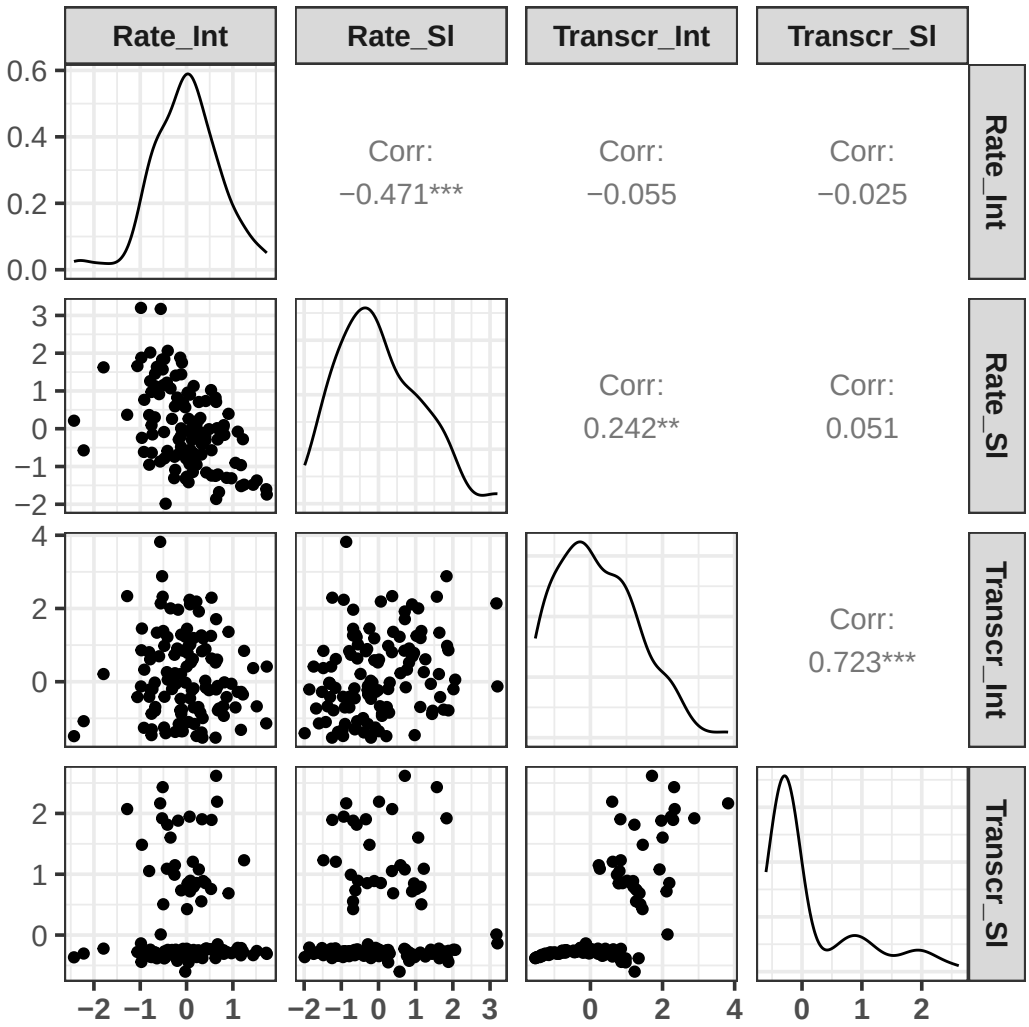

Figure 3.10: Pairwise correlations among the participant-level random intercepts ("Int") and slopes ("Sl") for both the rating ("Rate") and transcription ("Transcr") tasks.

$p = 0.008$). This indicates that participants with greater differentiation between guises in the rating task were more likely to transcribe regularized subject-verb agreement in the transcription task. No other terms were significant.

3.4 General Discussion

Across tasks, I found clear evidence that listeners across groups have some knowledge of variation in subject-verb agreement (SVA), and this knowledge affects their perception of ambiguous stimuli. I also found evidence that speakers of AAL—a variety with regularized SVA—are both more likely to expect regularized SVA in a neutral context and less likely to expect patterns that are ungrammatical in AAL.

Returning to the pre-registered hypotheses, all but two hypotheses were supported by the data. In Hypothesis 2C, I predicted that bidialectals "will rate sentences that are ungrammatical in both AAL and MAE as equally unlikely in both dialects." Instead, bidialectals rated ungrammatical sentences as more likely in AAL and MAE, though to a lesser degree than did monodialectals. As I described in Section 3.3.1.6, there are many reasons that this could be true, including the effects of standard language ideology on such a metalinguistic task. Additionally, in Hypothesis 2B, I predicted that bidialectals would "show a greater differentiation than monodialectals between MAE and AAL in perception," as measured by the transcription task. This did not emerge. Instead, there was simply a main effect of group, where bidialectal participants were more likely to transcribe regularized SVA in either guise, compared to monodialectal participants. This differs from the results for the rating task, where greater differentiation between guises was observed for the bidialectal participants, compared to the monodialectal participants.

One possibility here is that a true interaction effect exists, but there was not adequate statistical power to detect the effect. Though power simulations were a component of the pre-

registered participant numbers, many aspects of the model make statistical power low, since the effect required an interaction between groups in a logistic mixed-effects model with a complex random effects structure. Moreover, the proportions of regularized responses were all relatively low, and the reduced range of responses could reduce the ability to detect an effect.

Additionally, the two tasks were different in ways that might affect the ways in which participants draw upon their linguistic knowledge. The transcription task required participants to select between two possible resolutions of the ambiguity, while the rating task presented a given version of a sentence and asked participants to rate the likelihood of one presented form. This means that participants might have varied in the degree to which they were considering a non-MAE alternative in a given trial. This was partially by design, with the goal of separating recognition of a form from predicting a form in a neutral context. However, failure to transcribe regularized SVA does not necessarily constitute lack of knowledge of AAL. If in a listener's experience, AAL speakers can use both the regularized and MAE variant of a given variable, then both transcriptions are valid.

The two tasks also differed in the blocking choices for the two variables. Recall from Experiments 1 and 2A that there were apparent order effects when all items of a given guise were grouped together, with order of the guises counterbalanced across participants. Since the question of interest for this chapter was about differences between groups, I wanted to avoid an additional between-participants factor (guise order) that was not of theoretical interest. Due to the nature of the tasks, I resolved this in different ways. For the transcription task, it was possible to simply present the voices in random order, with both guises mixed together, instead of grouping items by guise. For the rating task, this approach would have made the task quite difficult, since participants would need to be explicitly prompted about the guise for each trial. Instead, items were blocked by guise, with all AAL items preceding all MAE items, as this order

had produced clearer guise effects in Experiment 1. While these approaches make independent sense for each task, it is possible that mixing the guises together weakens participants' ability to form differentiated expectations. Thus, it is possible that bidialectal participants do have the capacity to more strongly differentiate their grammatical predictions than do monodialectal participants, but they are more likely to do this when in contexts where only one dialect is being used. This might, in fact, be a more efficient processing strategy in a context where multiple dialects are being used, as it reduces the likelihood of prediction errors (see Valdés Kroff and Dussias, 2023, for discussion of this issue for bilinguals).

Regardless of these differences, these tasks provide converging evidence about participants' knowledge of different dialects' grammars. Almost all participants consistently *recognized* regularized subject-verb agreement in the rating task, and most participants were able to predict regularized subject-verb agreement at least some of the time in neutral contexts. Both tasks also provide clear evidence that listeners form a more granular understanding than "SVA is different in AAL, compared to MAE." In the rating task, we observed this as a condition effect, where participants in both groups gave higher likelihood ratings to the Grammatical-AAL sentences than the Ungrammatical-Agreement sentences, for the AAL guise. In the transcription task, we see this from the presence of regularized sentences, but the near-complete absence of "irregularization" in the filler sentences. These effects replicate what I observed in Chapter 2 with likely monodialectal speakers from Prolific. The present chapter expanded beyond that work by showing group differences across varying levels of exposure to AAL. Bidialectal participants had more granular knowledge, as measured by the rating task, and were more likely to predict regularized SVA, as measured by the transcription task.

Despite this, both groups maintained a broad preference for MAE: in the rating task, both groups gave higher likelihood ratings to ungrammatical sentences in the AAL guise than in the

MAE guise, and in the transcription task, participants predominately used the MAE variant, even if they used the regularized variant some of the time. As I discussed above, some of this may be attributed to standard language ideology, which generally favors MAE as the "correct" way to speak; thus, participants will be more likely to attribute descriptively ungrammatical utterances to the prescriptively "incorrect" guise, and when presented with the option to transcribe something in either a prescriptively correct or incorrect way, they choose the "correct" one.

Beyond this, many aspects of the task setting did not provide a supportive context for AAL. Though participants completed the study remotely and might have been located in a space where AAL is used, this study was associated with a university, which is a context where MAE is typically used. Moreover, the university in question was a predominantly White institution, and the instructions and communication around the study were conducted in MAE by White researchers.[9] Additionally, both tasks involved written language; in the rating task, items were written, and in the transcription task, participants provided written responses. While written norms for AAL have developed in the social media era (Blodgett et al., 2016; Jones, 2015), they differ from MAE writing conventions. All writing in the study used MAE conventions (and participants generally used these conventions in completing the transcription task), which might have made both tasks more natural for representing MAE than AAL. Since MAE forms were within all participants' linguistic repertoire, it makes sense that they would favor MAE forms given all of these factors.

[9]Since we did not interact in person and I never asked, I can't be sure how/if participants racialized me. Some email accounts might have shown an image of me. Both my name and the name Jan Edwards appeared on the consent form, neither of which are likely to be racialized as Black (though my surname is sometimes mistaken for an identically-spelled Arabic given name).

3.4.1 Issues of group definition

Defining participants' dialect backgrounds is a notoriously difficult problem. The one standardized assessment, the Diagnostic Evaluation of Language Variation – Screening Test (Seymour et al., 2003) is only normed for children (ages 4-12), so for adults, we must either rely on an intensive sociolinguistic interview or self-report. A sociolinguistic interview was not feasible for this study, and it is unclear whether it would be valid for the purpose of sorting participants into groups for a psycholinguistic study; a sociolinguistic interview is typically used for questions about the usage rates of particular sociolinguistic variables. Instead, I used self-report, which is also challenging, as participants may be reluctant to report usage of a stigmatized variety and might also have a different construct in mind for what constitutes a given dialect/language variety than the researcher. (This differs from bilingualism research; we might more reasonably expect a participant and a researcher to have aligned intuitions of what constitutes "English" vs. "Spanish" than "Mainstream American English" vs. "African American Language," even if research on translanguaging would question both dichotomies).

My design and results both present a path forward and raise some issues. The rating task takes about fifteen minutes (and could likely be shortened), and it shows clear group differences. This means that in future studies, it could be adapted and used as a secondary task to identify participants' experience of variation. It also sets up participants to inductively build a category for the lects of interest, making self-report categories be more meaningful.

However, the choice of a four-point scale proved problematic. While it was intended to inflect relatively coarse gradations for the purpose of defining two categories, I found that many White, university-based participants were reluctant to report that they *never* use AAL, responding with a "2" instead of "1" on the scale. This might reflect a reluctance on the part of some participants

to use the extreme of a scale, and unfortunately, on a four-point scale, the next option is adjacent to the middle of the scale. In the next chapter, I recruit from the same participant pool using a six-point scale and find that most White participants used 1 or 2. This supports the idea that the present results reflect minimal use of AAL, but not quite at the edge of a scale, rather than periodic use. (If 2/4 had reflected "near the middle," we would expect participants to select 3 on a six-point scale.)

It is an open question whether the difference between 1 and 2 on the AAL usage scale corresponds to a difference in Likert-type scale usage, a difference in behavior in the world, or a combination of both. The exploratory results by self-reported AAL usage were mixed. There are some reasons to believe that this pool of White participants would have substantial contact with AAL; in addition to the potential increasing appropriation of AAL in online spaces (see discussion in Holliday and Tano, 2021, and broader discourses on "Gen Z speak"), they are attending a university in a majority-Black county in the D.C. metro area, which is diverse along many dimensions. Despite persistent segregation and inequality, this increases the likelihood of meaningful exposure to AAL relative to less diverse regions.

For Black participants, increased knowledge of the diversity within AAL also might affect their approach to the task. As noted above in Section 3.3.1.6, the audio clips drew from a range of locations and speakers, with a focus on highlighting common features of AAL that are most different from MAE. To a White MAE speaker with limited exposure to African American communities, this might produce a broad notion of AAL—something that they know of but do not speak. To a speaker of a variety of AAL, this might lead to more conflicting ideas of factors such as region, class, and age. To a 20-year-old middle class AAL speaker from Prince George's County, Maryland, an example from a middle-aged lower-SES speaker from New York City will differ from their idiolect in many ways. However, the comparison to MAE throughout the study likely helped

to produce the intended intuitions. A broader limitation is that the focus on morphosyntactic features associated with more vernacular forms of AAL (i.e. African American Vernacular English) might not align with the middle class AAL that many of the college-educated participants likely speak, which tends to draw less upon the stereotyped morphosyntactic features of AAVE and more upon phonological, rhetorical, and lexical features (Weldon, 2022). Future work should examine a broader range of AAL speakers across the SES continuum, but the presence of an effect even for a convenience sample highlights the promise of such work.

It is also possible that participants' responses on the rating task, transcription task, and self-report are all tied to standard language ideology. Under this explanation, a participant with a stronger standard language ideology might be inclined to report lower usage of AAL, attribute any MAE-ungrammatical sentence to AAL on the rating task, and provide prescriptively "correct" answers on the rating task. This could weaken the inference that group differences in both tasks reflect differences in linguistic knowledge; instead, they might reflect different ideological stances at play in the metalinguistic aspects of the task. The present results cannot entirely exclude this explanation, but the transcription task does not provide a clean fit for this story. Participants did not have an issue transcribing the filler items with multiple negation, so it is not the case that participants merely "corrected" the speech that they perceived. Moreover, if participants knew that SVA is sometimes regularized in AAL and had the ideological stance that prescriptively incorrect forms are used in AAL, then we would expect *more* regularized transcriptions in the AAL guise. Instead, we find the opposite pattern, where lower self-reported AAL use is associated with *fewer* regularized transcriptions.

3.5 Conclusion

In this chapter, I compared two groups with different levels of experience with AAL in tasks designed to address (1) the granularity of their knowledge of regularized subject-verb agreement (SVA) when *recognizing* forms and (2) their ability to *predict* regularized forms in neutral contexts. On both measures, bidialectal participants showed greater knowledge of regularized SVA, but many monodialectal participants also showed some knowledge of SVA. Taken together, these findings indicate that listeners have mental models of others' grammars, but may vary both in the precision of the model and their propensity to draw upon it depending on task demands. Also, in both tasks, participants showed a general preference for MAE forms. This might partially reflect the offline nature of the measures and a tendency to favor MAE when providing a written response. In the next chapter, I will begin to address this limitation by using an on-line measure (eye-tracking in the visual world paradigm), which will allow us to see what interpretations are under consideration before participants give a final response—a measure that might be less biased toward MAE.

Chapter 4: Interpretation and processing of variation

So far, we have seen that listeners adjust their expectations regarding subject-verb agreement (SVA) depending on the language variety used by their interlocutor. There was widespread knowledge that regularized SVA is associated with AAL, and varying knowledge of the specifics of the SVA paradigm in AAL. Additionally, while participants showed a general preference for MAE forms, they were more likely to perceive ambiguous stimuli with regularized SVA when spoken by an AAL speaker, relative to an MAE speaker. In terms of the Labov (1973) framework, we have seen evidence of *recognition* of regularized SVA in supportive contexts, as well as *prediction* of regularized SVA in neutral contexts. However, the studies presented thus far do not offer any information about an additional component of Labov's framework: the *interpretation* of a particular form. This is because the truth-conditional meaning has not been affected by agreement morphology.

In this chapter, I turn to the interpretation of SVA morphology, again comparing monodialectal speakers of MAE to bidialectal speakers of AAL and MAE. Using the visual world paradigm, I study participants' incremental interpretations as they listen to sentences that might contain regularized SVA. In addition to testing interpretation—a different type of knowledge of variation—this chapter introduces questions of real-time processing, which have the potential to reveal differences in the time course in which knowledge of variation is deployed across different groups of listeners. I will begin with background information on both interpretation and on-line processing of variation, then present a study using the visual world paradigm, as well as a repli-

cation of the sentence rating task used in Experiments 1 and 3. To preview the results, I found parallel effects for both looking patterns and image selections, where listeners were more likely to look at and select an image that aligns with regularized SVA when the sentence was spoken by an AAL speaker, compared to an MAE speaker. Moreover, bidialectal participants were more likely than monodialectal participants to consider interpretations consistent with regularized SVA, but they did not differ in the degree to which they differentiated their interpretations depending on the dialect of the speaker.

4.1 Background

4.1.1 Interpretation

In most circumstances, SVA differences are unlikely to affect truth-conditional meaning; regardless of whether a participant perceives (39a), which aligns with MAE SVA, or (39b), which uses regularized SVA, only one interpretation is possible.

(39) a. He sit**s** still for hours.

b. He sit still for hours.

As Vaughn and Kendall (2018) note, this is broadly true of many sociolinguistic variables; for example, for (ING), it is quite rare for either possible realization to have an impact on word recognition, with only a small number of minimal pairs, like *robbin/robbing*, which can easily be distinguished in context.

However, there is still reason to believe that dialect differences can have an effect on interpretation in some circumstances. For example, Jones et al. (2019) found that court reporters often provided inaccurate transcriptions of AAL, shaped by a combination of phonological and morphosyntactic differences from MAE, and these court reporters also provided inaccurate para-

phrases of the sentences they had transcribed. Additionally, studies of lexical processing have found effects of dialect differences across a few paradigms, relying primarily on consonant cluster reduction, which is present in both MAE and AAL but more common in AAL. For example, Staum Casasanto (2009) presented forms like [mæs], which can be consistent with *mass*, as well as *mast* with consonant cluster reduction. Participants then selected a continuation that was compatible with either *mass* or *mast*, and they were faster to select *mast* interpretations when the sentence was paired with an image of a Black individual's face. Similarly, in a lexical priming paradigm, King and Sumner (2014) found that participants were more likely to be primed by words like *friendly* realized with [n] rather than [nd], when these forms were spoken by an AAL speaker.[1] There are also group-level differences in the usage of consonant clusters in lexical access. Edwards et al. (2014) compared the performance of AAL-speaking children across different levels of difference from MAE (dialect density) on a task where children selected between lexical items that differed only by a final consonant cluster, such as *goal* and *gold*. Children with a lower dialect density (less difference from MAE) were significantly more likely to use the absence of a final cluster (e.g., [goʊɬ]) to disambiguate between words (e.g., selecting an image of a *goal*, not *gold*).

There has been some evidence for similar findings in the area of morphosyntax, particularly SVA. Parallel to the findings by Edwards et al. (2014), Johnson et al. (2005) examined AAL-speaking children's usage of verbal *-s* to guide interpretation. They created materials where agreement morphology does lead to differences in interpretation. Children heard sentences like (40), where the verb started with /s/, meaning that a segmentation ambiguity made it impossible to determine whether there was a plural *-s* at the end of the noun. Because of this, only the presence or absence of verbal *-s* could be used to distinguish whether the subject of the sen-

[1] They use the term AAVE, intending to specify *vernacular* forms of AAL.

tence was singular or plural. Participants then selected between an image depicting the singular interpretation (one cat sleeping on a bed) or the plural interpretation (two cats sleeping on a bed).

(40) a. The cats sleep on the bed.

b. The cat sleep on the bed.

The authors found that AAL-speaking children did not use the presence or absence of verbal *-s* to select an image, meaning that they were at chance for these ambiguous items. Using a similar paradigm, de Villiers and Johnson (2007) compared AAL-speaking and MAE-speaking children from ages four through six. They found that MAE-speaking children start to reliably use the presence of verbal *-s* to interpret a subject as singular between ages five and six, while AAL-speaking children were at chance throughout the age range studied. Interestingly, for MAE-speaking children, this meant that they *produced* verbal *-s* according to the adult paradigm before using it to guide interpretation. Byrd et al. (2023) found similar results with the *was/were* alternation, indicating that this finding is robust to more acoustically salient morphological cues.

Issues with comprehension also can affect MAE speakers listening to AAL. As noted above, Jones et al. (2019) found general comprehension issues among court reporters listening to AAL, and a handful of studies have targeted specific morphosyntactic features of AAL. Beyer et al. (2015) found that AAL speakers, but not MAE speakers, correctly interpret stressed *BIN* (41a). The authors argue that this can constitute *pseudocomprehension*, where listeners believe that they have comprehended a sentence, even though they have accommodated the sentence with an incorrect meaning (e.g., 41b)

(41) a. Chad BIN using that cell phone. (Beyer et al., 2015)

'Chad has been using that cell phone for a long time'

b. Chad has been using that cell phone.

In other circumstances, comprehension is generally accurate across dialect differences. For example, Blanchette and Lukyanenko (2019b) found that MAE speakers accurately interpret sentences with negative concord even if they do not rate such sentences as grammatical, and Blanchette et al. (2023) found that MAE speakers correctly inferred the (wide) scope of negation in sentences like (42), even with ambiguous examples.

(42) Didn't everybody like the movie. (Blanchette et al., 2023)

a. Wide scope: Not everybody liked the movie. (correct interpretation)

b. Narrow scope: Nobody liked the movie. (incorrect interpretation)

While some of these studies (de Villiers and Johnson, 2007; Byrd et al., 2023) compared different groups of participants, none of these studies manipulated the speakers within a study, meaning that it is unknown whether participants will adjust their interpretations of SVA depending on who is speaking. To my knowledge, the closest parallel was an eye-tracking study by Lundquist and Vangsnes (2018), who tested how listeners used variation in grammatical gender in Norwegian in processing. They found a group of listeners who adjusted their use of gendered determiners depending on whether a given speaker makes a distinction between masculine and feminine grammatical gender. Ultimately, this does not yield a difference in off-line interpretation, since the temporary ambiguity associated with the determiner is resolved by the presentation of the noun.

4.1.2 Processing

The study by Lundquist and Vangsnes (2018) highlights an additional dimension of the psycholinguistics of variation that my studies have thus far left unaddressed: the time course in

which listeners deploy their knowledge of variation. Both the rating task and transcription task rely on off-line measures. Such measures could conceivably either overestimate or underestimate group differences in listening to SVA variation. On the one hand, if listeners tend to favor MAE in a university-based research setting, they might initially process items from an AAL speaker using AAL morphosyntactic rules, but then adjust their response to align with MAE grammar after incorporating metalinguistic judgments. On the other hand, listeners might initially generate incremental interpretations that align with their more dominant grammar, then at a later stage consider alternatives based on variation. For example, a monodialectal speaker of MAE might initially be guided by MAE grammar when listening to a speaker of AAL, then make adjustments at a later time. In any case, we might expect greater differences between monodialectal and bidialectal speakers in online measures, as bidialectal speakers are more likely to use their grammatical knowledge of AAL at earlier stages of processing than monodialectal speakers.

Previous studies have provided some indication that participants adjust their expectations with regard to *form* in online processing. For example, Squires (2019) used self-paced reading to study participants' processing of regularized SVA and found a reduced cost for regularized SVA when the forms were attributed to song lyrics, and Weissler (2021) found reduced P600 effects for the null copula when participants were listening to a Black AAL speaker, compared to a White MAE speaker. These studies were manipulating the speaker in the experiment, using university participant pools for listeners. Garcia et al. (2022) manipulated the listener groups and found that bidialectal participants did not show a P600 effect for regularized SVA, while monodialectal participants did show a P600 effect (the speaker in the experiment was a voice actor approximating Southern White English). While these studies do show differences in real-time processing of morphosyntactic variation, they do not show what *meaning* listeners derive from these forms. They are potentially compatible with an interpretation where listeners use a

"shallower" processing strategy in some contexts, not relying on agreement morphology, rather than generating an alternate interpretation of a given agreement cue.

4.1.3 Research questions and hypotheses

In this chapter, I ask how listeners use their knowledge of SVA variation between AAL and MAE to guide interpretation. Using eye-tracking in the visual world paradigm, I am able to ask this at both on-line and off-line levels:

1. Do listeners adjust their usage of SVA to make their *incremental interpretations* of whether a subject is singular or plural, depending on the dialect of the speaker and their own linguistic background?
2. Do listeners adjust their usage of SVA to make their *final interpretation* of whether a subject is singular or plural, depending on the dialect of the speaker and their own linguistic background?

For both on-line and off-line measures, if listeners adjust their interpretations according to the dialect they are hearing, we can predict that when participants are listening to AAL, they will be more likely to look at and select images whose interpretation is consistent with regularized SVA. Such an effect might vary by group (monodialectal vs. bidialectal) in a variety of ways. One possibility is that bidialectal participants will be more likely to differentiate their interpretations depending on their interlocutor, compared to monodialectals, since they have greater experience switching between different dialects. Alternatively, bidialectal participants might take a more global strategy, as we saw with the sentence transcription task, where they are generally more likely to consider a regularized interpretation than monodialectals, regardless of their interlocutor's dialect.

Beyond effects of speaker and listener, effects might be observed depending on the time

course of the measure. If the incremental looks and final image selections align, this would indicate that results are not driven by later-arriving metalinguistic judgments (see discussion in Campbell-Kibler, 2016). Conversely, if incremental looks indicate greater consideration of interpretations that are only available in AAL than can be observed for image selections, this would indicate that participants initially process sentences based on a favored grammar, then adjust their final response based on later-arriving metalinguistic information.[2]

4.2 Experiment 4: Methods

Participants listened to sentences spoken by both AAL and MAE speakers while viewing four images on a screen. They selected the image that best matched the sentence they had heard, and their eye movements were tracked throughout. This study has a 2x2 design: group, a between-participants factor (bidialectal speaker of AAL and MAE, monodialectal speaker of MAE); and the dialect of the speaker of the experiment, a within-participants factor.

4.2.1 Procedure

In the first part of the study, participants completed an eye-tracking task using the visual world paradigm. They heard a series of sentences (described in detail below) while four images were presented on a screen 1920x1080 BenQ XL24020T display. Audio was played through Bose noise-cancelling headphones. Eye movements were recorded by an EyeLink 1000 eye-tracker with a 500 Hz sampling rate. A five-point calibration was used (center, top, bottom, left, right), and all participants received a "good" validation score before the experiment began. To minimize trackloss, the audio file was triggered for each trial after the participant fixated to an animated

[2]While I have mostly considered metalinguistic explanations thus far, particularly in interpreting the results in previous chapters, there are logically possible alternatives to late effects being metalinguistic in nature. For example, effects might reflect competition between different linguistic forms, more akin to a garden path effect. This is an interesting area for future work, but since I do not observe any timecourse-based differences here, my results have limited bearing on such questions.

fixation point at the center of the screen for 400 ms. A drift correction was applied every 30 trials, and participants were re-calibrated as needed, such as when there was visible trackloss in general or in a particular part of the screen, or frequent issues with triggering the gaze-contingent start to the trial. I was the experimenter for all participants and sat beside the participant to operate the equipment. This might have affected the nature of the task, as I am a White monodialectal MAE speaker in my early 30s; even to the monodialectal participants, the presence of a non-peer researcher might have reduced the availability of non-mainstream forms.[3]

After the eye-tracking portion of the study, participants completed a series of tasks implemented in one Qualtrics survey. This portion was completed on a laptop in a separate room from the experimenter. Participants were asked open-ended questions about their experience with the eye-tracking portion of the study. They then completed an adapted version of the rating task from Chapters 2 and 3 (and described below), answered questions about their experience with each "Language Variety" in the study, answered demographic questions about themselves and their primary caregivers, and completed a language attitude questionnaire (from Filson, 2019). The entire procedure lasted approximately one hour.

4.2.2 Participants

Two groups of participants were recruited: White participants who were likely to be monodialectal speakers of MAE ($n = 44$) and Black participants who were likely to be bidialectal speakers of MAE and AAL ($n = 58$). Each group was targeted through a combination of email advertisements at the University of Maryland, a listing on the University of Maryland Credit SONA Psychology platform, and a listing on the University of Maryland Paid SONA Psychology Platform. Potential bidialectal participants were also recruited at Howard University, a historically

[3]I attempted to use more "informal" parts of my linguistic repertoire, particularly the *-in* variant and *gonna*, while interacting with the participants. This probably had minimal impact, but it could have slightly increased the availability of non-prestige forms in the artificial context of an in-lab experiment.

Black university. Participants completed the study in person on the campus of the college they attend (with the exception of one Howard University student who completed the study at the University of Maryland campus), and the same equipment was used at both locations. SONA listings were only displayed to participants in the pool if they indicated in a pre-screening survey that they are over 18 and speak English as their first language, with separate listings for participants who identified as "Non-Hispanic White, European-American" or "Black, Afro-Caribbean, African-American" (the labels designated in the system). Participants were compensated with either 1 course credit (reflecting one hour of participation) or a virtual gift card ($20 for monodialectal participants, $40 for bidialectal participants, to assist with the recruitment of a special population).

Near the end of the experimental session, participants completed a questionnaire where they self-reported their usage of and exposure to AAL on a 1-6 Likert-type scale. This was a change from the previous 1-4 scale, where a potential avoidance of the extremes of the scale required participants to provide answers that were adjacent to the middle of the scale. These responses are summarized by participant race in Figure 4.1. As we can see, White participants overwhelmingly indicated 1 or 2 for AAL usage, while Black participants showed a wider range, peaking at 4/6 (the upper half of the scale). As in the previous chapter, White participants reported exposure to AAL near the middle of the scale, while Black participants generally reported high exposure (mode at the maximum of the scale).

The monodialectal group ($n = 39$) included participants who identified as White (and no other race/ethnicity), and whose self-reported usage of AAL was 1 or 2 and self-reported usage of MAE was 5 or 6. The bidialectal group ($n = 36$) included participants who identified as Black and whose self-reported usage of AAL was 4-6 (the top half of the scale) and usage of MAE was 2 or higher.

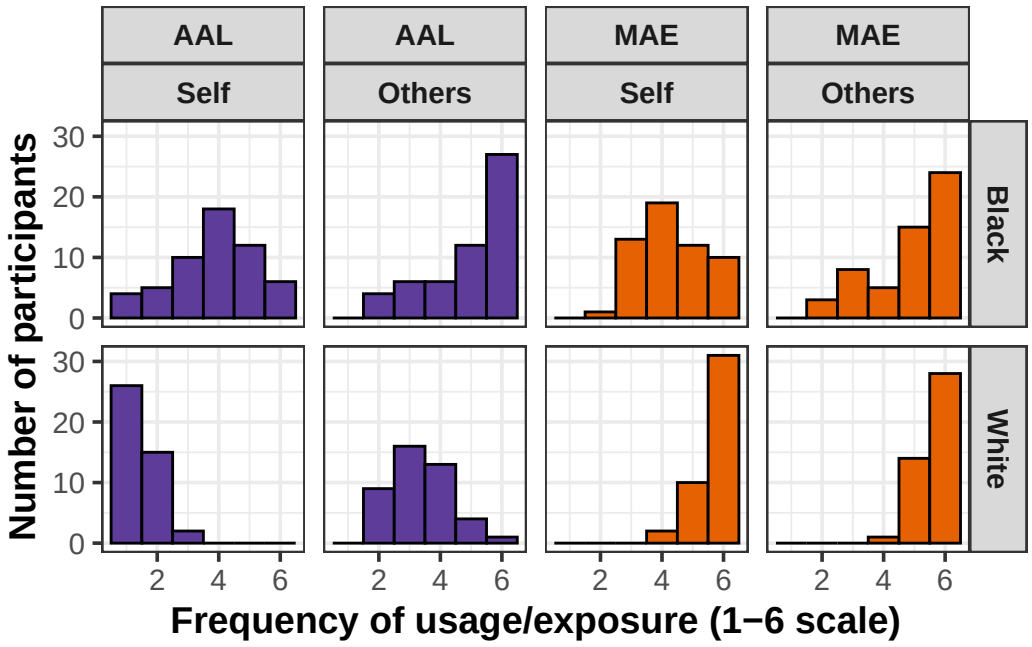

Figure 4.1: Particpants' self-reported usage of and exposure to both AAL and MAE on a six-point Likert-type scale. Participants are split by self-reported race, corresponding to the two recruited groups. They were subsequently selected for inclusion in the bidialectal and monodialectal groups based on their responses to self-reported usage questions.

A total of 5 participants were excluded from analysis due to equipment errors that resulted in restarting the experiment after it had begun. Additionally, 7 participants were excluded from analysis due to responding before the end of the sentence on more than 15% of the noun morphology and verb morphology trials (see below), as this likely indicates that they were monitoring for key nouns rather than listening to the sentence.

In the final analysis of the eye-tracking task, 34 participants were included in the bidialectal group, and 31 participants were included in the monodialectal group. Participants ranged in age from 18[4] to 24 (mean = 19.3). Most participants identified as female (n = 53), with the remaining participants identifying as male, nonbinary, multiple gender identities, or questioning. Participants' self-reported socioeconomic status (SES) on a ten-point scale (Adler et al., 2000) is visualized in Figure 4.2. Both groups averaged above the middle of the scale, with monodi-

[4]One participant wrote "12," but I take this to be a transposition error.

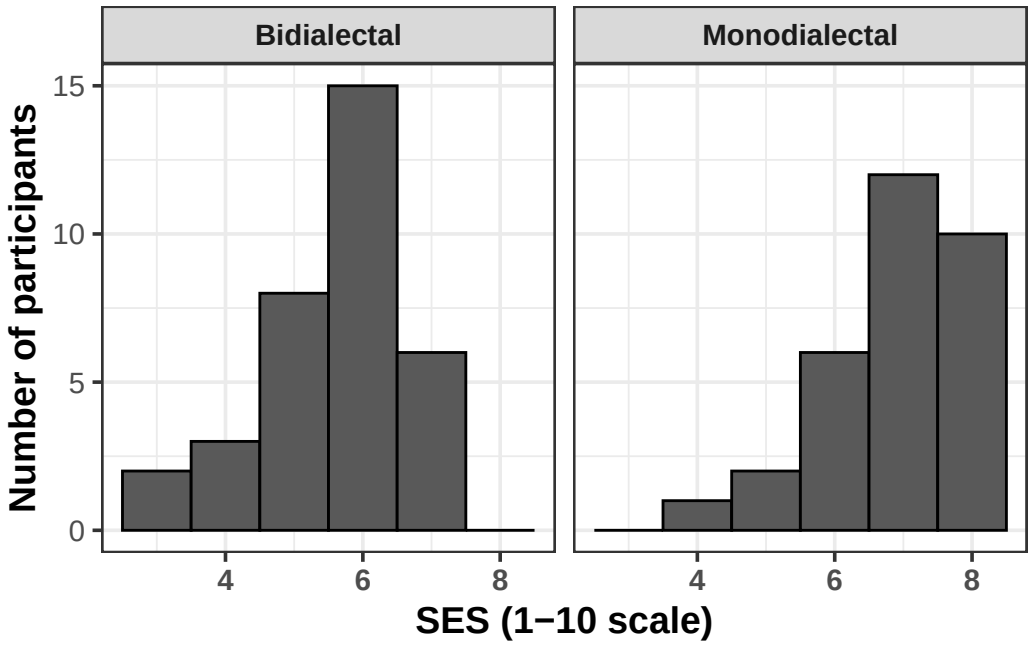

Figure 4.2: Histogram of self-reported socioeconomic status (SES) by group, for participants included in analysis.

alectal participants reporting a higher SES (mean=6.9, s.d.=1.04) than bidialectal participants (mean=5.59, s.d.=1.08). This difference between groups is statistically significant, based on a Welch two-sample t-test ($t(62.75 = -5)$, $p < 0.001$).

4.2.3 Materials

All sentences were recorded by four speakers: a male AAL speaker, a female AAL speaker, a male MAE speaker, and a female MAE speaker. These were the same individuals who recorded items for the sentence transcription task in Experiment 2C and Experiment 3, with the exception of the female AAL speaker. All speakers were current university students in their twenties and from the mid-Atlantic region of the United States. I will refer to the AAL vs. MAE distinction in the materials as the *guise* manipulation.

Experimental items were presented in eight blocks of 30 items. Each block contained only one voice, and the four voices were presented in random order in each half. To help establish

the grammar associated with each guise, the first two trials were always one negative concord and one null copula item (see below), followed by the remaining items in random order, with sequences of more than two of the same item type not permitted. Each participant received a unique randomized list. A full list of items is provided in Appendix C.

4.2.3.1 Critical verb morphology items

The purpose of the critical items was to force participants to use the agreement morphology on the verb to determine whether the subject of the sentence was singular or plural, adapted from the design used by Johnson et al. (2005). An illustrative example is *The duck(s) swim in the pond*. This type of item has two crucial components. First, the verb starts with /s/ and the noun ends in a voiceless stop, meaning that any plural *-s* would be co-articulated with the first segment of the verb. This means that a listener cannot use the number morphology on the noun to determine whether it is singular or plural. Second, the verb is in simple present tense form without an *-s*. In MAE, this unambiguously indicates that the subject is plural, while in AAL, this leaves the sentence ambiguous between singular and plural, since verbs can be zero-marked with both singular and plural subjects.

No verbal *-s* appeared in any of the experimental materials, so the experiment did not provide any indication that verbal *-s* was part of a given speaker's grammar. To make sure that the speakers across guises recorded equivalent sentences that were grammatical in the dialect they were representing, the noun was written with plural *-s* in the recording script (i.e., *The ducks swim in the pond*, not *The duck swim in the pond*). Speakers were instructed to co-articulate the relevant /s/-segments, and only versions of the items where they successfully did so were used.

Sentences were constructed using eight imagable nouns that end in voiceless stops (four humans, four animals): *cat*, *duck*, *rabbit*, *snake*, *acrobat*, *cook*, *infant*, and *pilot*. Four sentences

were created per item, yielding a total of 32 items, with a given verb used no more than twice across the list of items. All verbs started with /s/, and since this manipulation depends on a segmentation ambiguity, no verbs were used if removal of the /s/ would yield a different verb (e.g., *The cat(s) sleep* could be segmented as *The cats leap*).[5] Each participant heard a given sentence twice (once per guise and once per half of the experiment).

4.2.3.2 Noun morphology items

Since the manipulation requires participants to determine whether the subject of a sentence is singular or plural based on the absence of a monophonemic morpheme (/s/), one might be concerned that any guise effect reflects a general difficulty associated with using such a subtle cue, particularly in a less familiar dialect. To control for this, I included items where participants needed to use the presence or absence of plural *-s* on a noun—with neutral verb morphology—to determine whether the subject of a sentence was singular or plural. For these items, an auxiliary that does not change form depending on number agreement was used (43, 44). This allows for more direct comparison to the verb morphology items, since in both cases, the presence or absence of a single /s/ morpheme determines whether the subject of the sentence is singular or plural.

(43) The barber/barbers can gossip all day.

(44) The bear/bears did dent the trash can.

A total of 64 items were created, using the following nouns (four items per noun): *barber*, *carpenter*, *doctor*, *farmer*, *lawyer*, *pastor*, *soldier*, *teacher*, *bear*,*dog*, *bird*, *lion*, *monkey*, *squirrel*, *turtle*, and *zebra*. Each participant heard one version a given item. For a given noun, two items were presented in singular form, and two items were presented in plural form, with each voice represented once and each guise represented once in singular and once in plural.

[5] I am thankful to Sara Lowery for pointing out this issue.

4.2.3.3 Filler items

Filler items were used to establish the grammar of each guise. Two widely known phenomena were selected for this purpose: negative concord (45a) and null copula (46a). For the MAE guise, equivalent sentences were used with single negation and the negative polarity item *any* (45b), and with an overt contracted copula (46b). There were 48 items of each (24 per guise and 12 per voice).

(45) a. He **ain't** bring **no** jacket.

b. He **didn't** bring **any** jacket.

(46) a. She looking at the phone.

b. She**'s** looking at the phone.

For both sentence types, sentences were constructed to be dialect-neutral with regard to subject-verb agreement. This means that for the null copula items, only third person singular subjects were used, so that the MAE form would always be contracted *is* rather than contracted *are*. Additionally, as shown in (45a), for negative concord items using *didn't* in the MAE version, the AAL version used *ain't*. For null copula items, all items had a pronominal subject, as this is a more common environment for the null copula (Labov, 1969), and 32 of the items involved a verb ending in *-ing*, which is a common environment. While this was not systematically manipulated, the AAL speakers generally used the *-in* variant, and the MAE speakers generally used the *-ing* variant. There were also four items using a prepositional phrase (47) and eight items with an NP predicate (48), drawing from the human nouns used in the critical verb morphology and control noun morphology items. For nouns starting with a vowel, the AAL speakers sometimes elected to record with *a* instead of *an* (48), depending on what felt most natural.

(47) She on the phone.

(48) She a acrobat.

In addition to these guise-establishing filler items, there were 16 filler items for the nouns used in critical verb morphology trials. These followed the same structure as the noun morphology items, where the auxiliary verb does not differ depending on the subject's number (49). These items were always presented in singular form to reduce any bias toward plural for these nouns.

(49) The duck did quack so loud.

4.2.3.4 Images

The screen was divided into four quadrants, with one image in each quadrant. Clipart images were used; all images had a transparent background and filled the square area of interest at their larger dimension. For the verb morphology and noun morphology items, there was a singular target image (e.g., one duck), a plural target image (e.g., two ducks), a singular distractor image (e.g., one cat), and a plural distractor image (e.g., two cats). The plural version of a given image always involved both a replication of the singular image and a different image depicting that noun, but in the same style. When the target image involved a human, the distractor images were drawn from the other images of humans. When the target image involved an animal, the distractor images were drawn from the other animals. Other nouns that started with the same letter were excluded from the list of potential distractors to minimize potential cohort effects.[6] Otherwise, the selection of distractor for a given trial was random and independent of all other trials, and the assignment of each image type to a given quadrant was random within a trial.

For negative concord trials, the target image featured the object noun covered by a large red X. The alternatives were a foil image of the same noun without a red X, a distractor image

[6] I am thankful to Madison Lenhart for making this suggestion, as well as her considerable work preparing the materials for this experiment.

with a red X, and the same image without the X. For example, for *He ain't bring no jacket* (45a), a participant might see a jacket covered by an X (the target image), a jacket without an X, a book with an X, and a book without an X. For null copula trials, half of the items used the same types of images as negative concord trials (pairs of images with and without an X); this was to ensure that the presence of X's on the screen did not guarantee that an image with an X would be the correct answer. For the other half of the null copula items, participants saw images depicting four random singular nouns on the screen and selected the image that matched the word that appeared in the sentence.

4.2.3.5 Training

Participants completed eight training items before the start of the task. They were instructed to select the image that matches "some part of what the sentence is about." Each voice appeared twice in these training items, and each verb morphology target noun appeared once. The sentences were intended to be dialect-neutral with regard to morphosyntax and were designed to familiarize participants with the different types of picture selections they would make: differentiating between singular and plural (for verb morphology items and noun morphology items), selecting an image with an X for the negative concord (or single negation) items, and selecting the image of the object for null copula items. Participants received verbal feedback from the experimenter and asked clarifying questions as necessary.

4.2.4 Sentence rating task

After completing the eye-tracking task, participants completed an adapted version of the sentence rating task used in Experiments 1 and 3. Given the concern that the wide geographic range of speakers affected bidialectal participants' responses, only the voices from the experiment

and voices from the most recent D.C. corpus in CORAAL (Kendall and Farrington, 2021) were used for the guise-establishing audio. Additionally, null copula items were not included, since null copula had been used for the filler items in the preceding task.

Instead, to determine whether the effects observed in Experiment 3 extend to other phenomena, both well-formed (50a) and ill-formed (50b) habitual *be* items, taken from Wolfram (1982), were added to the task. These items were only included in the AAL guise, allowing me to use Wolfram's full item set (six pairs of items) with each participant, and providing comparable numbers to each grammaticality condition for the SVA items.

(50) a. Everytime I go there he be busy.

b. *I think he be busy today. (Wolfram, 1982)

Note that this task differs from Wolfram's. Wolfram used a forced-choice task, where items were presented in pairs. Sentences were presented with the word *is* instead of *be*, and participants selected the sentence where *be* would be more appropriate. Thus, the present study lets us examine whether comparable results can be yielded between Wolfram's task and the rating task that I have used in this dissertation.

4.3 Results

4.3.1 Image selection

4.3.1.1 Critical verb morphology items

Participants' image selections for the critical verb morphology items are provided in Figure 4.3. Trials were excluded from analysis if the participant gave a distractor response or responded before the end of the sentence. We can see that participants were more likely to select the singular image in the AAL guise than in the MAE guise and that bidialectal participants were

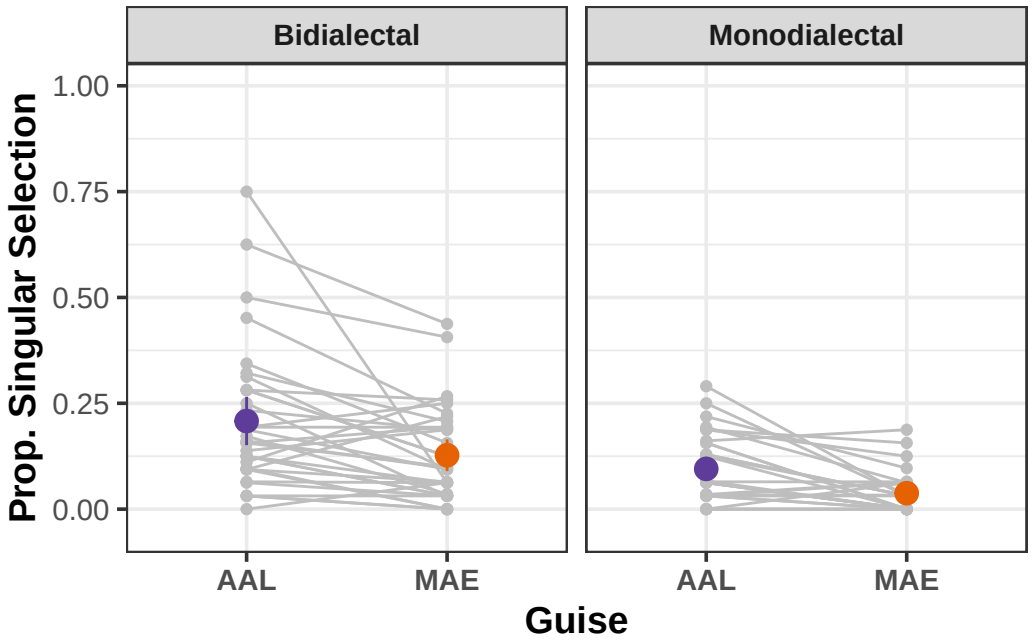

Figure 4.3: Proportion of items where participants selected a singular image in the critical verb morphology trials. Based on MAE SVA rules, only a plural interpretation is available, while for AAL, regularized SVA means that both a singular and plural interpretation are available. Light gray lines and points represent individual participants, while points in color represent means across participants. Error bars represent 95% confidence intervals, calculated as the standard error of partcipant-level means multiplied by 1.96.

overall more likely to select the singular image (compared to monodialectals), regardless of guise.

This was tested statistically using a logistic mixed-effects model.[7] The dependent variable was whether (or not) participants selected a singular image, with guise and group as effects-coded fixed effects (-0.5 for the AAL guise and the bidialectal group, 0.5 for the MAE guise and monodialectal group), as well as a guise-by-group interaction. There were participant- and item-level random intercepts, and guise-by-participant and guise-by-item random slopes. There was a significant intercept ($\hat{\beta} = -2.67$, $p < 0.001$), meaning that participants were significantly less likely to select the singular image than to select the plural image. There was a significant effect of group ($\hat{\beta} = -1.23$, $p < 0.001$), meaning that monodialectal participants were significantly less likely to select the singular image, compared to bidialectal participants. There was a significant

[7]Full regression tables for all models can be found in Appendix C.

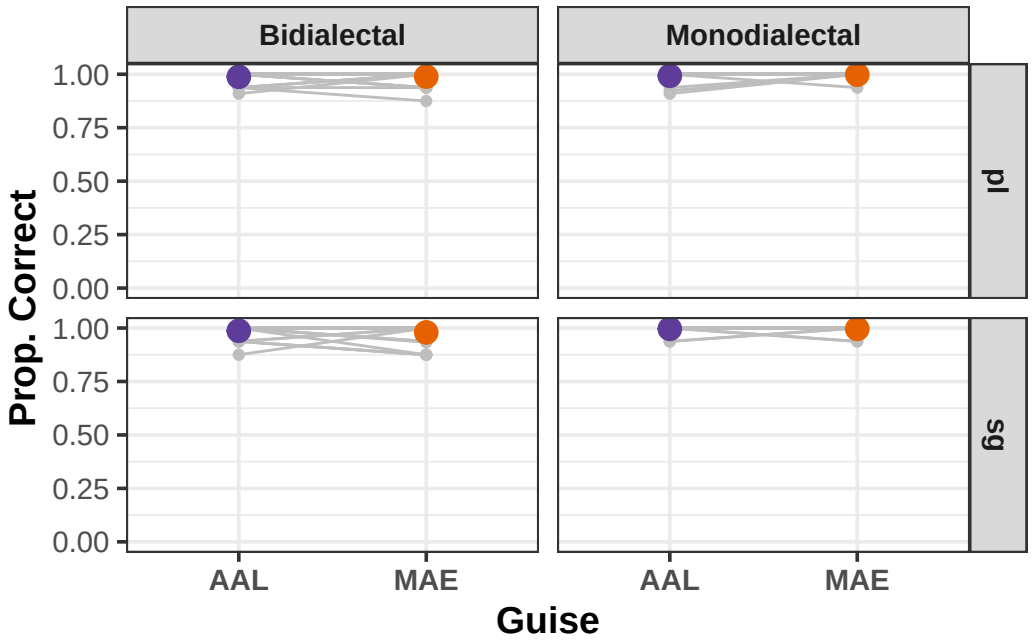

Figure 4.4: Proportion of items where participants selected the correct image in the control trials, based on the presence or absence of plural *-s* on the noun. There are no hypothesized effects based on dialect differences. Light gray lines and points represent individual participants, while points in color represent means across participants. Error bars represent 95% confidence intervals, calculated as the standard error of partcipant-level means multiplied by 1.96.

effect of guise ($\hat{\beta} = -0.81$, $p < 0.001$), meaning that participants were significantly less likely to select a singular image in the MAE guise, compared to the AAL guise. The interaction term was *not* significant ($\hat{\beta} = -0.38$, $p = 0.22$), meaning that there was no observed difference in the guise effect between groups.

4.3.1.2 Noun morphology items

Responses to the control noun morphology items are plotted in Figure 4.4, again excluding trials where a distractor image was selected. Overall, while accuracy is near ceiling for both groups, we can see that accuracy appears to be slightly lower in the bidialectal group, driven by a small number of participants who provided up to 3 incorrect responses (out of 64), as summarized in Table 4.1.

Table 4.1: Count of participants in each group, sorted by the number of incorrect responses for items relying on noun morphology.

Total Incorrect	Bidialectal	Monodialectal
0	17	23
1	9	8
2	3	0
3	5	0

This was tested statistically using a logistic mixed-effects model. The dependent variable was whether (or not) participants selected the correct image, with guise and group as effects-coded fixed effects, as well as a guise-by-group interaction. There were participant- and item-level random intercepts, and a guise-by-participant slope (the model failed to converge with guise-by-item random slopes).[8] There was a significant intercept ($\hat{\beta} = 5.35$, $p < 0.001$), meaning that participants selected the correct image more often than not. Contrary to my hypotheses, there was a significant effect of group ($\hat{\beta} = 1.23$, $p = 0.006$), meaning that bidialectal participants were significantly more likely to select an image that did not align with the noun's number morphology, compared to monodialectal participants. There was no significant guise effect ($\hat{\beta} = 0.66$, $p = 0.27$) or interaction ($\hat{\beta} = 0.75$, $p = 0.39$).

4.3.2 Looks

While the picture selection task provided preliminary evidence about the interpretations that participants considered, their image selections reflect the end of a process where multiple interpretations were potentially competing. Additional evidence comes from their gaze patterns, which can reflect interpretations that were incrementally under consideration as participants listened to the sentences. Areas of interest (AOIs) were coded for each of the four images on the

[8] Based on an automated model-building process using `buildmer`, the optimal model for this data does not have *any* random effects. I report the mixed-effects model because it allows for more direct comparison to the previous model, which is of primary interest. This does not have any effect on which regressors are significant.

screen, with looks to other regions (e.g., the fixation point at the center of the screen) and track-loss grouped together as non-looks. Data were collapsed into 20 ms time bins, examining the 1220 ms window following the offset of the verb for verb morphology items and following the offset of the noun for noun morphology items (shifting the window forward 150 ms to account for saccade planning). Since I did not have specific hypotheses regarding the timing of any effects, this window was selected to be maximally long without extending beyond the coverage of the data; only 50.2% of trials extended beyond this range, since the median picture response time following the verb (including the 150 ms adjustment) was 1221.

4.3.2.1 Verb morphology items

Looks to each AOI are plotted in Figure 4.5. For the research questions, the relevant comparison is looks to the singular target image, which is available in AAL, compared to looks to the plural target image, which is the only available interpretation in MAE and also an available interpretation in AAL. This comparison is plotted in Figure 4.6, which shows looks to the singular target image relative to total looks to either the singular or plural target image. Parallel to the picture selection data, there is an apparent difference between guises, where there is a higher proportion of looks to the singular image in the AAL guise, relative to the MAE guise, regardless of participant group, and there is also an apparent difference between groups, where bidialectal participants had a higher proportion of looks to the singular image. There also appears to be a greater difference between guises for the bidialectal participants toward the beginning of the time window and a greater differences between guises for monodialectal participants toward the end of the time window.

These differences were evaluated statistically using a cluster-based permutation analysis (Maris and Oostenveld, 2007). This approach to analysis allows us to test for time windows

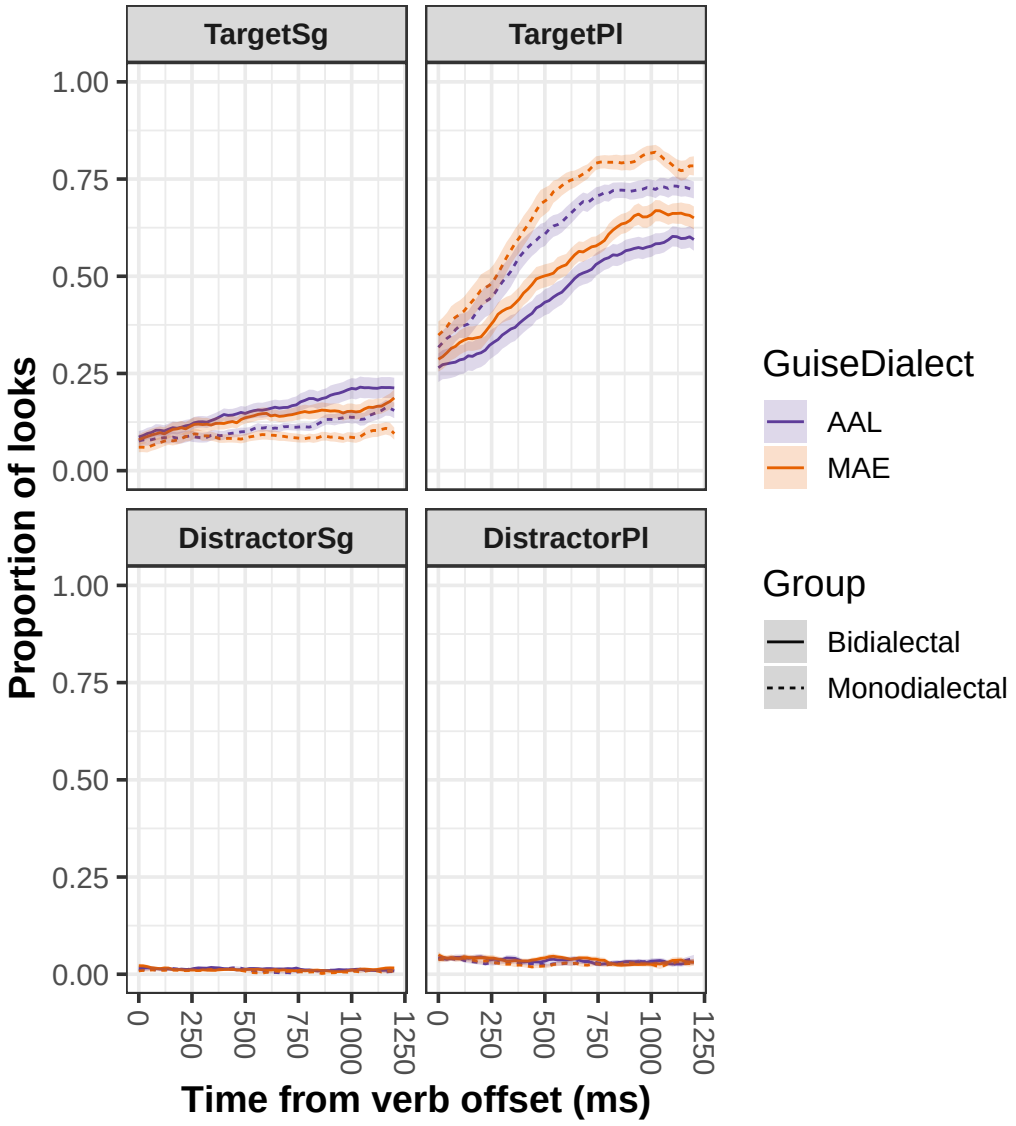

Figure 4.5: Looks to each area of interest for the critical verb morphology sentences. Looks are averaged across 20 ms bins, and error bar represents that standard error, calculated based on participant-level means. Non-AOI looks are included in the denominator.

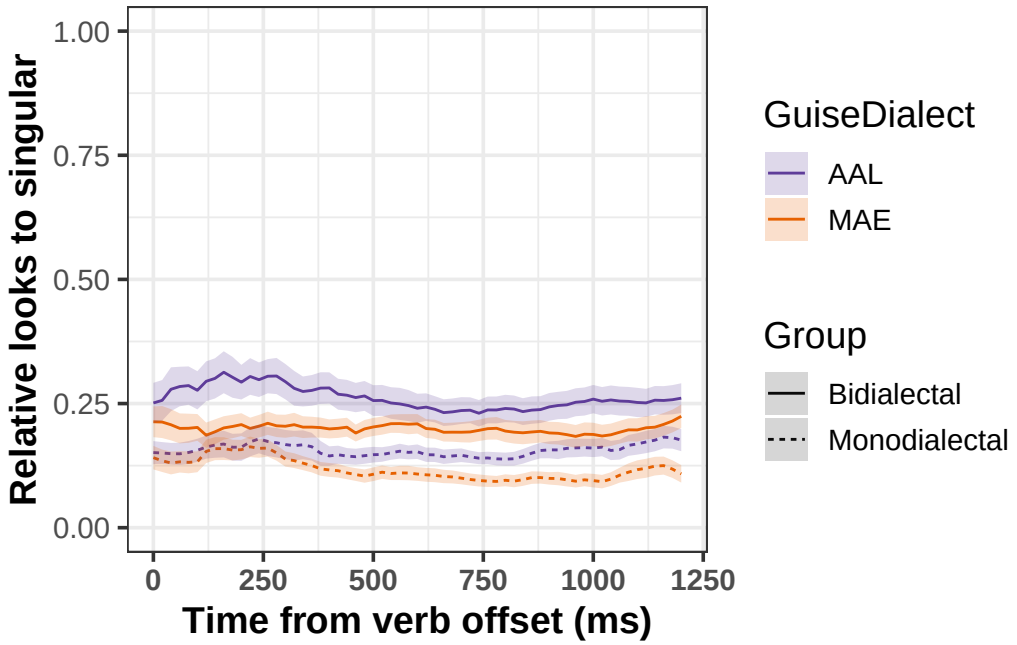

Figure 4.6: Looks to the singular target image, relative to total looks to the singular and plural target images. Looks are averaged across 20 ms bins, and error bar represents that standard error, calculated based on participant-level means.

where a given effect is significant, while accounting for concerns about multiple comparisons and researcher degrees of freedom in the selection of the relevant time window. This technique is common practice for EEG data and a good alternative to other curve-fitting approaches that have been used for eye-tracking data, due to recent concerns that temporal autocorrelation leads to anticonservative results (Huang and Snedeker, 2020). Following the procedure outlined by Maris and Oostenveld (2007), this analysis involved two steps. In the first step, I calculated time windows where there was a significant effect in the same direction for all consecutive time bins. To do this, I fitted a logistic model for every time bin, where the dependent variable was the number of samples with a look to singular, relative to the number of samples with a look to plural, and independent variables were group and guise (both effects coded) and their interaction. (Random effects were not included in this model, as non-independence for participants and items was accounted for in the simulation procedure.) I extracted the z statistic for each term in the model, and a given window was interpreted as significant if the absolute value of the z statistic

was 1.96 or higher. The size of a cluster—a type of area under the curve measure—was calculated as the sum of absolute values of consecutive significant z statistics with the same sign.

In the second step, I simulated the largest possible cluster for each term under the null hypothesis for the data. Each simulated data set preserved the key sources of nonindependence in the data: participants and items. For a given simulated data set, participants were drawn with replacement from the pool of participants included in the analysis and randomly assigned to a group (bidialectal or monodialectal), preserving the totals in each group. Since each item appeared once per guise per participant, the guise labels for a given item were randomly flipped within each participant, with each item having a 50% chance of having its guise labels reversed. As with the true data set, trials were excluded if a participant responded before the end of the sentence, but if the other version of the item was included, it had a 50% probability of having its guise label changed. For each simulated data set, I used the same procedure to estimate the largest significant cluster for each term, based on a logistic model in each 20 ms time bin. This process was repeated 10,000 times, creating a null distribution for the largest cluster for each term.

For both the main effect of guise and main effect of group, the largest cluster in the true data spanned the entire interval, meaning that each effect had a (negative) z statistic with an absolute value greater than 1.96 for every time bin. The simulated set of clusters can be considered to be the null distribution for each parameter, and p values were calculated as the proportion of simulated clusters that were larger than the largest cluster observed in the data. The effect was significantly different from the simulated distributions for both guise ($p < 0.001$) and group ($p < 0.001$). The largest cluster for the guise-by-group interaction spanned from 820 to 1040 ms, but it was not significant ($p = 0.82$).

4.3.2.2 Noun morphology items

Picture selections for the control noun morphology items are plotted in Figure 4.7. We can see that participants generally show increased looks to the correct image over the entire time window, with overall higher looks to the correct image from monodialectal participants, with no clear effects of guise.

For appropriate comparison to the verb morphology items, looks are plotted in Figure 4.8 as a proportion of looks to the correct image relative to looks to either image corresponding to the noun, regardless of number morphology. In addition to showing the apparent group effect, there is an initial preference for the plural item, meaning that when hearing a word like *ducks*, participants immediately looked to the plural image, but when hearing a word like *duck*, participants would often initially look to the image of two ducks before looking at the image of one duck.

As with the verb morphology items, a cluster-based permutation analysis was performed to determine the statistical significant of these effects. The procedure was equivalent to the one used for verb morphology items, adjusting to account for the fact that items were not repeated between guises for the noun morphology items. Instead, a given noun appeared twice per guise, once in plural form (with a nominal *-s*, *ducks*) and once in singular form (*duck*). This means that for a given noun and subject number, there is one trial per guise, and the guise labels for these trials were randomly reversed in each simulated data set. Due to their different pattern, singular and plural items were analyzed separately. The effect of group was significant for singular trials in the 180-1220 ms range ($p < 0.001$) and for plural trials in the 20-1220ms range ($p < 0.001$). The effect of guise was not significant for singular ($p = 0.16$) or plural ($p = 0.39$), and the interaction term was not significant for singular ($p = 0.48$) or plural ($p = 0.73$).

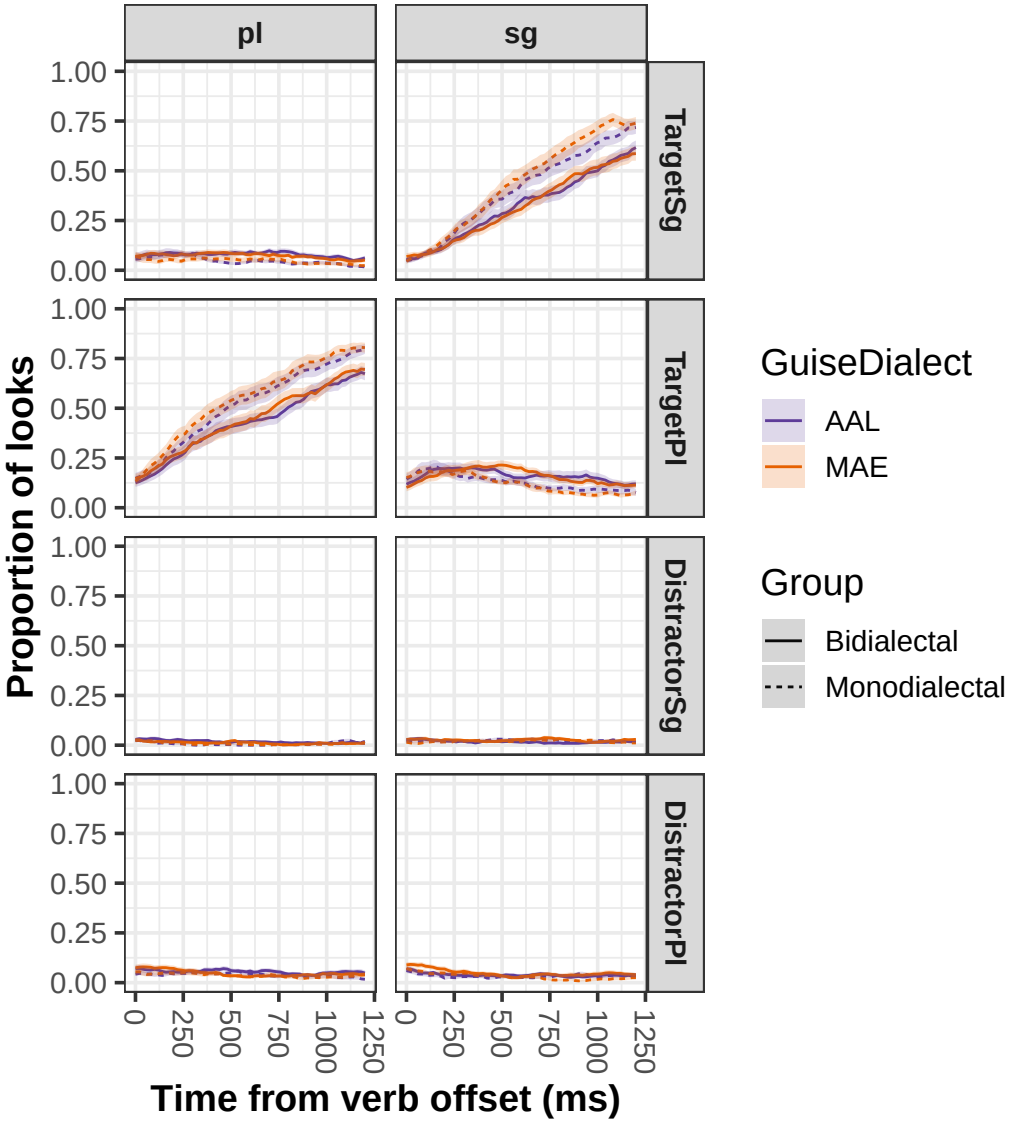

Figure 4.7: Looks to each area of interest for the control noun morphology sentences, split by singular vs. plural trials (e.g. *duck* vs. *ducks*). Looks are averaged across 20 ms bins, and error bars represent standard error, calculated based on participant-level means. Non-AOI looks are included in the denominator.

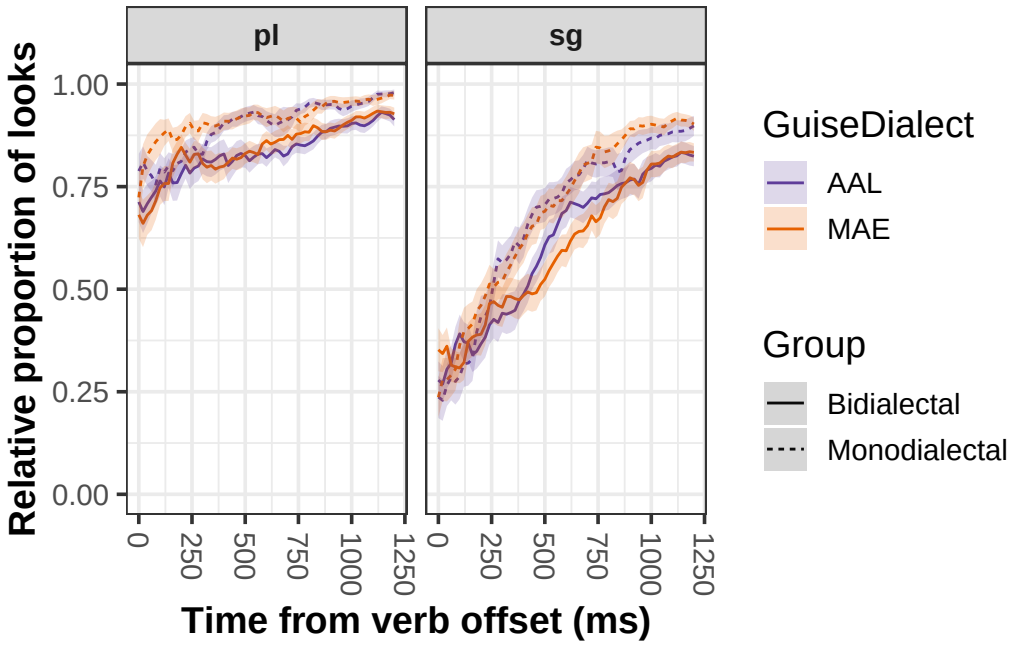

Figure 4.8: Looks to the correct image, relative to the correct noun with the wrong morphology, for the control noun morphology sentences. Looks are averaged across 20 ms bins, and error bar represents that standard error, calculated based on participant-level means.

4.3.3 Reaction times

This study was not designed for analysis of reaction time (RT) data; participants were instructed to respond after the end of the sentence, but they received sufficient information to respond at variable times before the end of the sentence (after the verb for verb morphology items and the noun for noun morphology items). However, given the group difference in the noun morphology items, the RT data provides useful information to understand participants' general approach to the task. For example, if a participant is using a "shallow" processing approach and monitoring for key lexical items to quickly move through the task, we might expect faster RTs, while if they perceive an item to be ambiguous or had an attentional lapse, we might expect slower RTs. Thus, I analyzed reaction times from the sentence offset, excluding items where participants responded before the end of the sentence. Items were also excluded if the log-transformed RT was more than three standard deviations from the mean across participants.

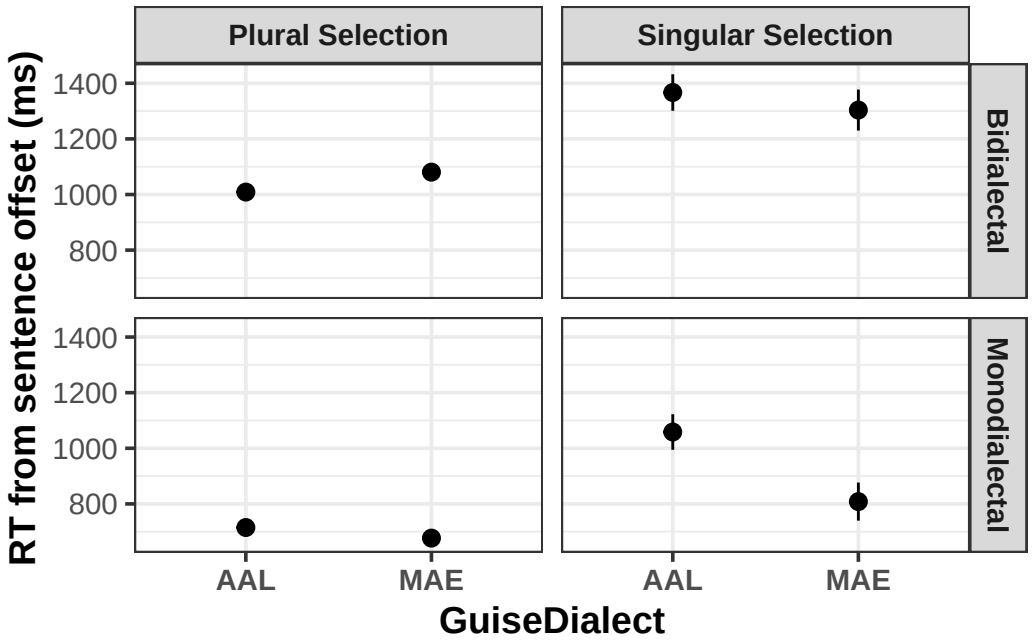

Figure 4.9: Reaction times (from sentence offset) for verb morphology items. Error bars represent the standard error, ignoring participant- and item-level nonindependence.

RTs for the verb morphology items are plotted in Figure 4.9. Bidialectal participants had overall slower RTs than monodialectal participants, and RTs were generally slower when participants ultimately selected the singular image, which is only compatible with AAL grammar. There is also an apparent difference between guises for monodialectal participants when selecting a singular image, where RTs were slower in the AAL guise, compared to the MAE guise.

This was confirmed using a linear mixed-effects model, with log-transformed RT as the dependent variable; effects-coded fixed effects of guise, group, and response (singular vs. plural); all two- and three-way interactions; item- and participant-level random intercepts; and guise-by-item and guise-by-participant random slopes. There was a significant main effect of group ($\hat{\beta}^{\star} = 0.27$, $p = 0.001$), meaning that bidialectal participants had longer RTs than monodialectal participants. There was a significant main effect of response type ($\hat{\beta}^{\star} = 0.1$, $p < 0.001$), meaning that there were longer RTs for singular, compared to plural picture selections. There was no significant main effect of guise ($\hat{\beta}^{\star} = 0$, $p = 0.25$), meaning that there were no differences in RT

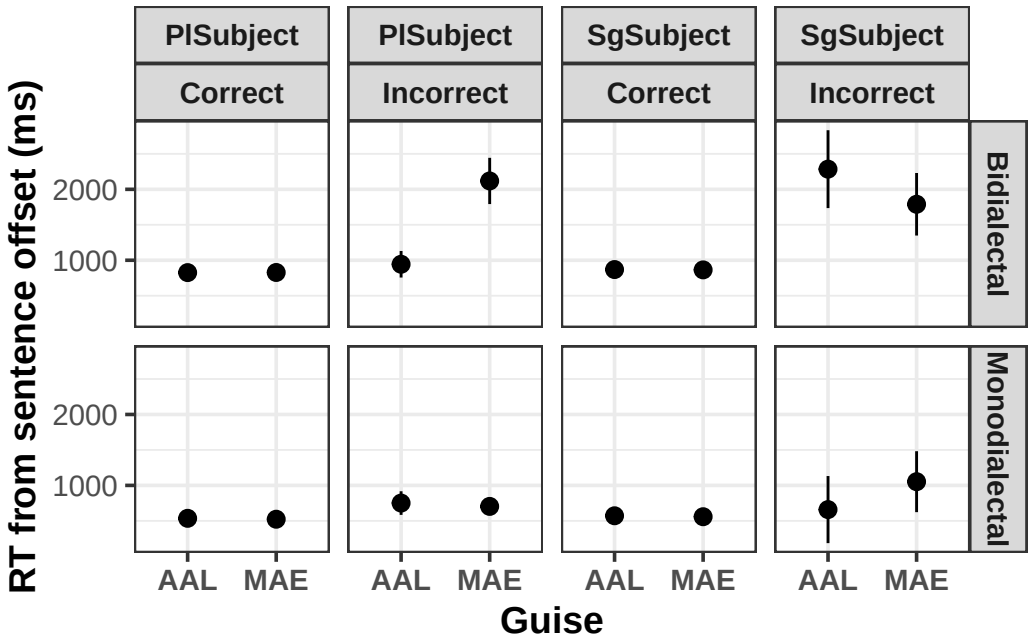

Figure 4.10: Reaction times (from sentence offset) for noun morphology items. Error bars represent the standard error, ignoring participant- and item-level nonindependence.

between guises when averaging across groups and response types. The group-by-response type interaction was significant ($\hat{\beta}^{\star} = -0.05$, $p = 0.004$), the guise-by-group ($\hat{\beta}^{\star} = -0.04$, $p = 0.07$) and guise-by-response type interactions were marginal ($\hat{\beta}^{\star} = 0.03$, $p = 0.1$), and the three-way interaction was not significant ($\hat{\beta}^{\star} = -0.01$, $p = 0.71$).

RTs for noun morphology items are plotted in Figure 4.10. Again, we see slower RTs for bidialectal participants than for monodialectal participants, and there are slower responses for incorrect responses, compared to correct responses. There are possible guise effects within a few incorrect facets, though given the small number of trials that this reflects, this should be interpreted with caution.

These patterns were evaluated statistically using a linear mixed-effects model, with log-transformed RT as the dependent variable; effects-coded fixed effects of guise, group, and response (correct vs, incorrect); all two- and three-way interactions; and item- and participant-level random intercepts. The addition of subject number (singular vs. plural), along with all possible interaction

terms, did not significantly improve data-model fit ($X^2(8) = 9.71$, $p = 0.29$), so it was not included in the model. There was a significant main effect of group ($\hat{\beta}^\star = -0.28$, $p < 0.001$), meaning that bidialectal participants had longer RTs than monodialectal participants. There was a significant main effect of response correctness ($\hat{\beta}^\star = 0.06$, $p < 0.001$), meaning that there were longer RTs for incorrect responses, compared to correct responses. There was no significant main effect of guise ($\hat{\beta}^\star = 0$, $p = 0.44$), meaning that there were no differences in RT between guises when averaging across groups and response types. There were no significant interaction terms.

In summary, for all item types, we can see that bidialectal participants generally had longer response times, and response times for incorrect responses were longer than for correct responses. This suggests that the accuracy effects are not the result of a speed/accuracy tradeoff that varies between groups. Instead, they reflect trials where participants had increased uncertainty or difficulty.

4.3.4 Sentence rating task

4.3.4.1 Subject-verb agreement items

Results for the sentence rating task are provided in Figure 4.11; 1 bidialectal participant and 7 monodialectal participants are excluded for incorrect answers to a comprehension question for the task instructions.

These results were evaluated statistically using a linear mixed effects model, with response (shifted to 0 as the center of the scale) as the dependent variable; fixed effects of grammaticality condition, guise, and group (dummy coded); all two- and three-way interactions; participant- and item-level random intercepts; and random slopes for grammaticality condition and guise for both participants and items. For ease of interpretation, I analyze pairwise comparisons; the full regression table and a list of estimated marginal means can be found in Appendix C. There

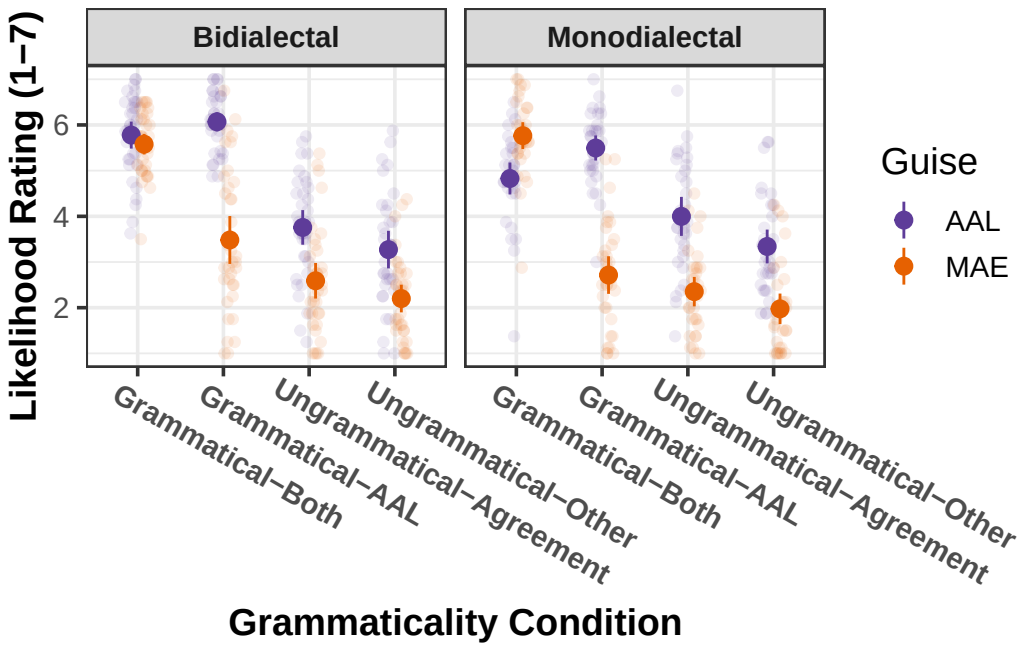

Figure 4.11: Participant ratings for both subject-verb agreement item types in each guise and grammaticality condition. Semi-transparent points represent participant-level means, and error bars represent 95% confidence intervals, calculated based on participant-level means.

were significant guise effects within a given group-grammaticality pairing (ps< 0.001), with one exception: there is no significant difference between guises in the Grammatical-Both condition for bidialectal participants (which corresponds to grammaticality in the world). For monodialectal participants, this difference was significant and negative, meaning that they gave lower likelihood ratings in the AAL guise, compared to the MAE guise. For all other grammaticality conditions, the AAL guise received higher likelihood ratings than the MAE guise.

Additionally, the difference between adjacent grammaticality conditions was significant for most comparisons. For bidialectal participants, the difference between Grammatical-Both and Grammatical-AAL was not significant in the AAL guise, and the difference between Ungrammatical-Agreement and Ungrammatical-Other was not significant in the MAE guise. All other conditions showed decreasing likelihood ratings moving across the levels (ps< 0.01). For monodialectal participants, there was no significant difference in the MAE guise between the Grammatical-AAL and Ungrammatical-Agreement conditions or between the Ungrammatical-Agreement condition

and Ungrammatical-Other conditions; this aligns with the fact that all of these sentence types are ungrammatical in MAE. All other conditions showed decreasing likelihood ratings across the levels (ps< 0.001), with the exception of the difference between Grammatical-AAL and Grammatical-Both in the AAL guise, where the difference was in the opposite direction ($p < 0.001$).

To serve as a parallel to the analysis for Experiment 3, I directly compared the groups in their ratings on the Grammatical-AAL and Ungrammatical-Agreement items, as this provides a comparison of the granularity of their knowledge. I used the same random effects structure as above, but fixed effects were effects-coded to allow for interpretation as main effects. There was a significant effect of grammaticality condition ($\hat{\beta} = -1.27$, $p < 0.001$), indicating that participants gave higher ratings in the Grammatical-AAL condition, relative to the Ungrammatical-Agreement condition. There was a significant effect of guise ($\hat{\beta} = -2.05$, $p < 0.001$), indicating that participants gave higher ratings in the AAL guise, relative to the MAE guise. There was a marginal effect of group ($\hat{\beta} = -0.33$, $p = 0.06$), indicating that bidialectal participants potentially gave overall higher ratings, compared to monodialectal participants. There was a significant condition-by-guise interaction ($\hat{\beta} = 1.28$, $p < 0.001$), indicating that the difference between grammaticality conditions was greater in the AAL guise, relative to the MAE guise. There was a significant condition-by-group interaction ($\hat{\beta} = 0.67$, $p < 0.001$), indicating that bidialectal participants showed an overall greater difference between the grammaticality conditions than monodialectal participants, regardless of guise. There was no significant guise-by-group interaction ($\hat{\beta} = -0.33$, $p = 0.33$) or three-way interaction ($\hat{\beta} = -0.31$, $p = 0.16$).

These findings partially replicated the patterns found in Experiment 3. The bidialectal participants showed greater differentiation between the Grammatical-AAL and Ungrammatical-Agreement conditions, compared to the monodialectals, and monodialectal participants indicated that Grammatical-Both items were less likely in AAL than MAE, while bidialectal participants

gave equivalent ratings to both guises for the Grammatical-Both items. Taken together, these patterns suggest a greater understanding of the systematicity of AAL in the bidialectal participants, compared to monodialectal participants. However, as before, bidialectal participants gave higher ratings to ungrammatical items when they were attributed to AAL, perhaps reflecting standard language ideology.

There were also some differences from the results found in Experiment 3. First, monodialectals showed slightly greater sensitivity to grammatical patterns of AAL, giving lower likelihood ratings for Ungrammatical-Agreement items than Grammatical-Both items. We also see that bidialectal participants gave higher baseline ratings for regularized SVA in the MAE guise, compared to monodialectal participants. That is, they view regularized SVA as broadly more likely. This aligns with the interpretation data.

4.3.4.2 Habitual *be*

For the AAL guise only, participants also rated both well- and ill-formed items using habitual *be*, taken from Wolfram (1982). These items are plotted in Figure 4.12. As we can see, bidialectal participants overwhelmingly differentiated between the well-formed and ill-formed uses of habitual *be* (91.43% gave higher likelihood ratings for correct than incorrect items), while monodialectal participants did not (56.25% gave higher ratings for correct items, indicating near chance performance).

These patterns were confirmed using a linear mixed-effects model. Group and condition were dummy-coded, with the monodialectal group and incorrect condition as the reference levels. There were participant- and item-level random intercepts, condition-by-item and condition-by-participant random slopes, and a group-by-item random slope. There was no significant effect of grammaticality condition ($\hat{\beta}^{\star} = 0.1$, $p = 0.46$), meaning that there was no difference between the

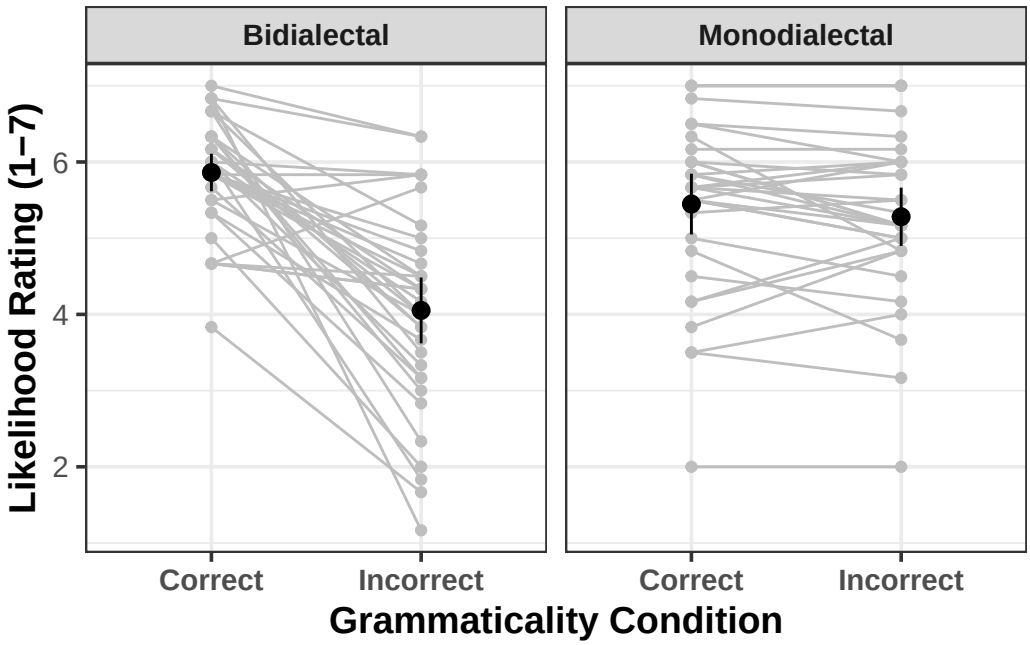

Figure 4.12: Ratings for habitual *be* items (Wolfram, 1982). Each line represents one participant, and the large dark points represent grand means, with error bars representing 95% confidence intervals based on participant-level means.

incorrect and correct items for monodialectals. There was a significant group effect ($\hat{\beta}^{\star} = -0.75$, $p < 0.001$), meaning that bidialectal participants gave significantly lower ratings for incorrect items than did monodialectals. There was a significant interaction ($\hat{\beta}^{\star} = 1$, $p < 0.001$), meaning that the difference between grammaticality conditions was greater for bidialectals than for monodialectals. *Post hoc* pairwise comparisons using `emmeans` confirm that there is a significant difference between conditions for the bidialectal group ($p < 0.001$) but not the monodialectal group ($p = 0.88$).

4.4 Discussion

In this study, I asked how participants adjust their *interpretation* in both on-line and off-line processing depending on the dialect of their interlocutor and their own linguistic experience. Participants heard sentences like *The duck(s) swim in the pond*, where the subject is unambiguously plural in MAE, but ambiguous between singular and plural in AAL. In both the online and

off-line data, we saw that a singular interpretation was more likely in an AAL guise (spoken by an AAL speaker) and when the participant was bidialectal (heard by an AAL speaker). Thus, participants did shift their interpretation depending on who they were listening to, as well as their broad experience with regularized-subject verb agreement. However, there was no statistically significant interaction, meaning that in this task, bidialectal listeners did not differentiate their interpretations based on their interlocutor more than monodialectal listeners.

4.4.1 Effect of group for noun morphology trials

The group effect in the noun morphology trials was surprising, since this was a control condition where no difference was predicted.[9] This weakens the inference that the group effect for noun morphology trials reflected something specific to their interpretation of agreement morphology on a verb. (In the case of the guise effect, there was no effect for either group for noun morphology items, so the effect cannot be attributed to issues of intelligibility in a given guise.)

There are several reasons why this might have occurred. One possibility is that there actually is some variation in plural marking in AAL. Zero-marked plural on nouns has sometimes been described as a feature of AAL, particularly among children (e.g., Washington and Craig, 1994), but this is contested for the adult grammar; if there is some apparent absence of plural marking, it is likely limited to nouns of measure (e.g., *fifty cent*) or consonant clusters (Green, 2002; Labov, 2012). However, if this were driving the effect for bidialectal participants, we would expect a difference between singular and plural items; zero-marked (singular) nouns might be interpreted as plural, but *-s*-marked nouns would not be interpreted as singular. No such difference exists in the data.

[9] In a secondary analysis that I did not include above, I randomly sampled a subset of participants such that the groups were balanced in their picture selections for the noun morphology items. Effects remained the same for the verb morphology items, and surprisingly, the looking patterns remained the same for the noun morphology items (increased looks to the correct item in the monodialectal group).

Another possibility is that bidialectal participants differed from monodialectal participants on some general aspect of task performance, like their attentiveness to the task; perhaps they were slightly more likely to listen for key words rather than fully parsing each sentence. This could have occurred due to some uncontrolled factor, or it could be that their inclination toward a different task strategy reflects a group difference that is relevant to the research questions at hand. For example, perhaps the frequent shifting between dialects promoted a shallower processing strategy, and bidialectal participants' increased sensitivity to dialect-shifting led them to be more likely to use such a strategy than monodialectal participants. However, the reaction time data show that bidialectal participants took more time than monodialectals to respond to each item. Thus, if the bidialectal participants used a "shallower" processing strategy, this did not result in "speeding through" the task. If anything, monodialectals were more prone to rush, as evidenced by their higher exclusion rate for responding before the end of a sentence and the higher proportion of monodialectals who incorrectly answered the comprehension question for the instructions to the rating task. (This was not a difficult task, since they could re-read the instructions before responding.)

Additionally, the somewhat incongruous context for AAL usage might have been broadly distracting for bidialectal participants in a way that did not affect monodialectal participants. That is, bidialectal participants are likely accustomed to AAL in predominantly Black social contexts, while the White monodialectal participants are never in such contexts, rendering the task novel, but not odd. Furthermore, given the relationship among race, language, and power, the presence of a White MAE-speaking experimenter in a study about language could have triggered stereotype threat (Steele and Aronson, 1995) in some Black participants, leading to general issues with task performance.

4.4.1.1 Lack of interaction

These results do not support the idea that bidialectal participants show more fine-tuned adaptations to their interlocutor's grammar; if this were the case, we would expect a significant interaction between guise and group. While such an interaction appears to be present in Figure 4.6, this effect was not significant based on the simulation procedure. Thus, while we can conclude that bidialectal participants adjust their on-line interpretations, we cannot make any conclusions about differences from monodialectals.

As with any null effect, there are essentially three types of plausible interpretations: no true effect; a true effect, but low statistical power; or a theoretically true effect in the world that was incorrectly operationalized in the experimental design. First, bidialectal listeners might not be strongly inclined to adjust their interpretations based on the grammar of their interlocutor. As I noted in the introduction to this chapter, it is relatively rare that two different dialects will generate an identical string with different truth conditions. Thus, most of the time, it probably makes sense to interpret a given form according to the available rules in a listener's repertoire. For example, if a bidialectal listener hears a sentence with habitual *be* uttered by someone who otherwise uses MAE, they might default to the habitual interpretation. Thus, while both groups did show some knowledge of the interpretations available for *The duck(s) swim in the pond* in AAL, extra experience with dialect shifting in the world does not lead to more dialect-contingent differentiation in interpretation.

Second, there might be an interaction effect in the world, but statistical power was too low to detect this effect. Power is particularly low for statistical interactions, and eye-tracking data has considerable noise, paired with uncertainty in the field regarding appropriate statistics. Additionally, the proportions of looks to the singular image were near zero, so floor effects might

have reduced the ability to detect *any* effect, particularly subtle ones. Taken together, these factors make it plausible (if not likely) that we would fail to observe a true statistical interaction, even with a relatively large sample size. Moreover, there was a numerical difference in the magnitude of the guise effect between groups, where bidialectal participants showed an earlier effect and monodialectal participants showed a later effect, raising the possibility that an interaction might be found with a considerably larger number of participants or items.

Finally, it might be the case that bidialectal people differentiate their interpretations depending on their interlocutor's grammar to a greater degree than do monodialectals, but some aspect of this experiment did not elicit this. As I previously mentioned, the experimental setting was not particularly supportive for AAL use, both due to the presence of the experimenter and the location in a university lab. While the parallels are inexact, recent research on bilingualism amplifies this concern; in a bilingual code-switching study using EEG methods, Kaan et al. (2020) found that the presence of other speakers can have an impact on processing. Additionally, the experimental items used individual sentences designed according to a template and without any context. While this a common practice in psycholinguistic studies, such decontextualized sentences could be particularly problematic for studying socially-mediated sentence processing.

Relatedly, the study switched back and forth between guises. While the blocked design partially mitigates this issue, the presence of speakers of both varieties within the task might lead participants to use a more dialect-general processing strategy, rather than tailoring their expectations to a given dialect. This limitation was perhaps amplified by the fact that participants never received direct information about a given speaker's usage of regularized SVA. Although this makes it particularly noteworthy that a guise effect was observed, a much larger effect might have been observed if participants were able to combine their prior knowledge of AAL SVA rules with specific knowledge of a particular speaker's repertoire.

4.4.2 Other limitations

Another potential issue is the selection of the bidialectal group, which had a largely middle class background. As Weldon (2022) notes, middle class AAL often relies more on phonological (particularly prosodic) and rhetorical features of AAL, along with some lexical items, rather than the morphosyntactic features associated with early descriptions focusing on African American *Vernacular* English. Thus, when these participants report that they use AAL, they might be thinking of other features (even though morphosyntactic features were included in all of the examples on the questionnaire). This concern is partially mitigated by their performance on the sentence rating task, particularly their sensitivity to the grammatical rules of habitual *be*, which was not observed for monodialectal participants. Regardless, it is possible that these participants still favor MAE-compatible grammatical forms, but others who use AAVE grammar would show different effects.

Another limitation of the study is the apparent baseline plural bias for the items. For the verb morphology items, this might have been influenced by the fact that the speakers in the recording session were actually producing plural nouns, which means there might have been subtle phonetic cues favoring a plural interpretation. However, the plural bias was also observed for unambiguously singular items in the noun morphology control sentences, where this issue was not present. This probably does not represent an underlying bias in the world toward a plural interpretation for these items; based on a Google ngram search, most of the nouns used in the study appear more frequently in singular than plural form (with the exception of *bird*, *acrobat*, and *farmer*, which are more frequently plural, and *teacher*, which is neutral). This leaves two more likely explanations. First, the plural images were inherently more visually complex than the singular images, which might have made them more visually salient, regardless of participants'

linguistic representations. Second, if participants favor a plural interpretation for the verb morphology items, the experiment is slightly biased toward plural answers, so participants might have become sensitive to the distributions of the experiment.

Given the plural bias, revision might be necessary for participants to arrive at a singular interpretation. In the case of the unambiguously singular noun in the noun morphology controls, participants did this successfully. However, in the case of the verb morphology sentences, there was no need for revision; since a plural interpretation is available in AAL, participants could make their picture selections according to their initial bias. If they initially fixated on the plural image, the rest of the sentence remained consistent with the fixated image. Future work might overcome this by forcing a singular interpretation with subsequent material in the sentence, but such an approach would differ substantially from the present design, as it would require that participants receive direct evidence in the experiment that a given speaker uses regularized SVA.

4.4.3 Conclusions

Despite these limitations, it is clear from this study that listeners can adjust their interpretations of morphosyntactic variables depending on their interlocutor's grammar. Indeed, it is remarkable that this was observable, given the general bias toward plural, likely stemming from a combination of image salience and the favoring of MAE forms. Moreover, these differences in interpretation can be observed as listeners build incremental interpretations of a sentence, indicating that the results observed thus far are not simply a result of off-line metalinguistic reasoning. Instead, listeners use their model of a speaker's grammar as they listen to a sentence. Given the limitations of the experimental setting, future work should see whether the observed effects are amplified in a more ecologically valid context, understanding what factors participants use to guide their real-time interpretations.

Chapter 5: General Discussion

In this dissertation, I have described a series of experiments designed to answer the following question: what do listeners know about morphosyntactic variation, and how does this knowledge guide perception, interpretation, and on-line processing? I have focused on differences in subject-verb agreement (SVA) in African American Language (AAL) and Mainstream American English (MAE), comparing participants across levels of experience with these varieties. Across tasks, I found clear evidence that most listeners have some knowledge of regularized SVA, and they are more likely to attribute regularized SVA to speakers of AAL, compared to speakers of MAE. There were also differences between groups, the details of which varied by experiment.

This concluding chapter is organized as follows. First, I review the key findings across the different experiments, noting both areas of convergence and divergence across tasks, as well as limitations across tasks. Second, I consider more broadly what constitutes "knowledge" of variation, with an eye toward elaborating the Labov (1973) approach. Third, I review how these findings bear on key questions and models in both psycholinguistics and sociolinguistics. I conclude with a discussion of open questions and ways that future work might address them.

5.1 Review of key findings

Throughout this dissertation, I have used three different tasks to test different types of knowledge of variation, as described by Labov (1973). The sentence rating task (a novel task) was designed to elicit grammaticality judgments for both AAL and MAE and to test *recognition*

knowledge of SVA patterns; the transcription task (adapted from Buson et al., 2018) was designed to test participants' *predictions* of different forms in a neutral context; and the visual world paradigm task was designed to test participants' *interpretations* of regularized SVA. In Chapter 2, I developed the rating (Experiment 1) and transcription (Experiment 2) tasks, using a national sample of likely MAE speakers recruited through Prolific; in Experiment 3, I compared bidialectal and monodialectal participants' responses on the rating and transcription tasks. In Experiment 4, I compared bidialectal and monodialectal participants' responses on the eye-tracking task and the rating task. To synthesize the findings, I will discuss the results task-by-task.

5.1.1 Recognition knowledge: Rating task

In the sentence rating task, participants heard naturalistic examples of AAL and MAE, then rated how likely a speaker of each variety would be to say a given sentence. Across groups and versions of the task, participants reliably attributed regularized SVA (*They was nice*) to AAL but not to MAE. In Labov's terms, this shows that participants *recognize* regularized SVA. It also suggests that participants *evaluate* the social meaning, since they attributed regularized SVA to one dialect and not another. I should note, however, that the social meaning was not directly quantified in this task; I did not, for example, ask participants to rate how "educated" or "friendly" a given sentence sounded. Instead, my primary measures focused on linguistic categories. Regardless, participants reliably racialized the guises in their open-ended responses, so they had some indexical representation for regularized SVA. The design of the rating task also allows for direct probing of the *granularity* of participants' knowledge of SVA variation. In the AAL guise, participants rated regularized SVA as more likely than uncommon forms that are not attested in either variety (*He were nice*). The difference between the ratings for regularized vs. uncommon forms was greater for bidialectal participants, compared to monodialectal participants,

indicating greater granularity of knowledge of AAL's SVA paradigm.

These findings of *recognition*, *evaluation*, and *granularity* all align with usage patterns in the world. However, there were additional reliable effects where participants' responses diverged from usage patterns in the world. Across groups, monodialectal participants showed a guise effect for sentences that are grammatical in both AAL and MAE, such as *He was nice*; such sentences were rated as less likely in the AAL guise than the MAE guise, and sentences with regularized SVA were rated as more likely in the AAL guise, compared to sentences that are grammatical regardless of dialect. Bidialectal participants showed no such difference, giving equally high likelihood ratings to sentences that are grammatical in the variety in question. We can think of this effect as the mirror image of the group difference for unattested forms like *He were nice*; monodialectal speakers have a clear association between regularized SVA and AAL, but an unclear representation of the rest of the agreement paradigm in AAL. In contrast, bidialectal participants view any AAL-grammatical form as grammatical in AAL, regardless of its status in MAE.

However, both groups shared one area of divergence from patterns in the world: they rated sentences that are ungrammatical in both AAL and MAE—even sentences with phrase structure errors like *She was eat*—as more likely in AAL than in MAE. If participants were rating according to actual usage, they would have given equally low ratings to both groups. As discussed in Chapter 3, there are a variety of factors that might explain this, including the influence of standard language ideology (Lippi-Green, 1997) on a metalinguistic task, where participants might hold the belief that AAL is "incorrect" language, rendering *any* MAE-ungrammatical form more likely in AAL than in MAE.

5.1.2 Prediction in neutral contexts: Transcription task

While the granularity of even MAE speakers' ratings shows that most U.S. English speakers associate regularized SVA with AAL, participants completing the rating task had the advantage of seeing the regularized form and then determining whether it seems familiar. As Labov's framework implies, a stronger test of knowledge is whether a listener can *predict* whether a given variant will appear, given a neutral context. The transcription task was designed to test *prediction* knowledge. (The psycholinguist reader is reminded here that this term might have different associations than Labov intended; while Labov's approach implies top-down aspects of processing, it need not require that anything occur in advance of the relevant form.)

In the transcription task, participants heard sentences where SVA was acoustically ambiguous, and they transcribed these sentences. In the guise manipulation, sentences were either spoken by an AAL speaker or an MAE speaker, with the guise reinforced by filler items using negative concord (and in some cases, null copula). Crucially, this task constitutes *prediction* knowledge, since the acoustic signal was equivalent across guises, and it was always consistent with a string that is licensed by the grammar of MAE; MAE-speaking participants had no need to diverge from their dominant dialect (in critical items) to complete the task. This task also does not require metalinguistic judgment, instead focusing on how listeners' perception of equivalent signals changes depending on their mental model of a speaker's grammar.

Across versions of the task, there was a significant guise effect, where participants were more likely to make regularized transcriptions in the AAL guise, compared to the MAE guise. There were also some situations that shifted the baseline rate of regularized transcriptions, meaning that there were also more regularized transcriptions in the MAE guise. Of primary theoretical interest, bidialectal participants were more likely to make regularized transcriptions, compared

to monodialectal participants. This suggests that regularized SVA is broadly more available to bidialectal participants, at least in a task where both dialects are present in the context. In Chapter 2, we also saw that regularized response rates can be adjusted depending on different aspects of the task design. For example, participants were more likely to give regularized responses in the MAE guise if it followed the AAL guise than if it preceded the AAL guise. We can interpret this similarly: the presence of AAL in the context has made regularized responses more broadly available; when MAE comes first, there has not yet been a violation of the default status of MAE in an experimental context, so participants essentially do not consider possibilities that differ from MAE in an otherwise ambiguous context.

5.1.3 Interpretation: Visual world paradigm

In Chapter 4, I asked whether participants use their knowledge of SVA variation to guide the *interpretation* of verb agreement morphology. The absence of verbal *-s* provided a unique test case: in MAE, a zero-marked simple present tense verb cannot have a third person singular subject, while in AAL, any subject can appear with a zero-marked verb. Thus, in situations where the verb's agreement morphology is the only source of information about whether the subject of the sentence is singular or plural, a listener might consider multiple interpretations when listening to AAL but only one interpretation when listening to MAE. While there was still an overwhelming preference for the interpretation that is compatible with MAE, I found that participants were more likely to consider a singular interpretation when listening to the AAL guise, compared to the MAE guise. There was a group difference, where bidialectal participants were more likely to consider an interpretation that aligns with regularized SVA, compared to monodialectal participants. As with the transcription task, there was no evidence that bidialectal participants adjusted their expectations to a greater degree than monodialectal participants. This task also allowed us to

test how participants' usage of dialect-contingent SVA cues unfolds over time. The off-line picture selections and on-line looking patterns generally aligned, with effects observed across the entire analysis window. This suggests that the adjustment of linguistic expectations occurs from the earliest stages in processing, rather than reflecting later metalinguistic judgments.

5.1.4 Relationships among different types of knowledge

At the individual level, I found some evidence of cross-task correlations. In particular, participants who showed the greatest level of granularity for the AAL guise in the rating task were more likely to transcribe regularized forms in the transcription task. In other words, participants who gave high ratings to *They was happy* and low ratings to *He were happy* in the AAL guise would be more likely to transcribe *He sit still*.

There were also consistent group effects across tasks: bidialectal participants demonstrated an increased expectation that regularized SVA occurs. This was most evident in the group effects for both the transcription and visual world paradigm tasks, but it also can be observed for the rating task; bidialectal participants gave higher likelihood ratings in the MAE guise for sentences with regularized SVA, relative to ungrammatical agreement in the MAE guise, and this difference was less reliable for monodialectal participants. This suggests that bidialectal listeners have a shifted baseline in their expectations of regularized SVA, relative to monodialectal listeners. Beyond this, as described above, both groups showed a general MAE bias, which likely reflects the opportunity to draw on standard language ideology in completing each task. For ambiguous items in the transcription and visual world paradigm tasks, this meant that they favored the option that aligns with the grammar of MAE, and for the rating task, this means that they rated ungrammatical items as more likely in AAL than MAE. Thus, measures of recognition, prediction, and interpretation generally align, suggesting that the types of knowledge outlined by

Labov (1973) do probe shared representations.

We might be tempted to ask whether there is a hierarchy of knowledge: for example, does successful interpretation of a given variant entail that a given person also can accurately predict the morphosyntactic environments in which that variant appears? There were large effects for the rating task and increasingly subtle effects for the transcription and visual world paradigm tasks. However, while it is reasonable that other type of knowledge entail recognition as a baseline, we cannot make strong inferences about the relationships among the relevant constructions from the size of effects in the tasks used here. The rating task differed substantially from the other two tasks. Rating allowed for a (relatively) continuous measure and used unambiguous items, while the transcription and eye-tracking tasks relied on ambiguity and dichotomous dependent variables. The combined ambiguity and need for logistic models reduces the ability to detect underlying differences. Thus, larger effects in the rating task do not necessarily entail that a greater degree of knowledge was shown on that task. (On the other hand, the rating task had higher metalinguistic demands than the rating and transcription tasks, but this did not seem to impede participants' performance.)

Beyond this, differences between the tasks extend beyond the different "type of knowledge" they were testing, weakening our ability to consider the relationships among different types of knowledge. In particular, each task took a different approach in the degree to which the guises were blocked vs. mixed together. At the "blocked" extreme, for the rating task, all AAL guise items preceded all MAE guise items (though in Chapters 3 and 4, this task always followed another task that featured both AAL and MAE). At the "mixed" extreme, the transcription task in Chapter 3 presented items in fully random order, so the guise could alternate between AAL and MAE from trial to trial. The eye-tracking task split the balance between the two, with eight blocks presented in random order, meaning that there were fewer occasions to switch between

guises, and these always occurred at clear break points. The pilot studies in Chapter 2 make it clear that the broad availability of AAL in the context affects how participants respond to MAE items, so some differences between the tasks could be attributed to how much switching occurred throughout a task. Future work will be necessary to make stronger claims about any implicational hierarchy in types of knowledge.

5.1.5 General limitations

Across the experiments presented in this dissertation, there are several common limitations that are worth reiterating here as we consider the results in the aggregate. Perhaps the greatest concern is the artificial nature of the experimental context, which might have made AAL feel broadly infelicitous to participants. Academic contexts generally favor MAE, particularly when White individuals (such as me!) are present. Moreover, the speakers and items were presented without context. In the case of the rating task and transcription task, the sentences were at least drawn from naturalistic speech, but in the eye-tracking task, sentences were artificial and followed a template to allow for time-aligned analyses. While the sentences in the rating and transcription tasks might have been more naturalistic, these tasks had a different disadvantage relative to the eye-tracking task: they involved written representations of AAL, using writing conventions that differ from those used for AAL. Future work should consider creative task designs to make the context more supportive of AAL use, recognizing that doing so might sacrifice some degree of the level of control that is typically favored in experimental studies.

An additional area where the design is not quite representative of the world is the selection of participant groups. The presence of group effects suggests that I was able to identify meaningfully different groups, but I am limited by the self-reported nature of the data. Most crucially, I have no direct evidence that participants ever use regularized SVA, and it is possible that self-reported

AAL usage reflects non-morphosyntactic features. For the group comparison, I am limited by the use of a university population, which represents a minority of the U.S. population: according to the National Center for Education Statistics, 37% of Black 18- to 24-year olds and 38% of White 18- to 24-year olds were enrolled in college in 2021. While this subset might be appropriate for studying a monodialectal MAE-speaking experience, the more middle class skew of the sample might not represent the experience of many Black speakers of AAL. This likely means that effects of group are underestimated in the data, and it is possible that with other populations of AAL speakers, the predicted group-by-guise interactions would have been observed.

Finally, while I have described the two guises throughout the experiment as "AAL" and "MAE," more work is necessary to determine what those categories mean to participants. The guises differed along multiple dimensions. This dissertation focused on the linguistic differences, particularly the presence of non-mainstream grammatical variants, although sociophonetic differences were also present. Beyond this, multiple social categories might have been under consideration, such as the racialization of each speaker, whether the speaker is more generally minoritized, and inferences about the regional background of the speaker (while all speakers were from the mid-Atlantic region, some participants identified AAL with the South, and many commented on likely "urban" origins of AAL speakers). I have begun to address this issue with follow-up studies, and preliminary results suggest that the presence of non-mainstream grammatical variants is crucial for the effects observed throughout the study. Future work in this area will be important to further link these findings to the sociolinguistic literature on social evaluation and the sociolinguistic monitor.

5.2 Revisiting "knowledge"

Throughout this dissertation, I have discussed different types of "knowledge" of variation, appealing to a framework outlined by Labov (1973). Given the present findings, to what extent is his approach a helpful carving of the hypothesis space, and to what extent is some modification or elaboration required?

As a reminder, Labov identified a series of six "abilities" to test the degree to which someone has "assimilated" a grammatical rule of a different dialect: (1) "recognizing" that a particular variant is grammatical to some people, (2) "evaluating" the social meaning, (3) "interpreting" the (truth-conditional) meaning, (4) "labeling" the meaning in isolated contexts, (5) "predicting" the syntactic and semantic contexts in which a variant may appear, and (6) "using" the form in context. In these terms, participants *recognized* regularized SVA in the rating task and *evaluated* that it is associated with AAL in both the rating task and the observed guise effects in all of the tasks. They *interpreted* the meaning of regularized SVA (in some cases) in the eye-tracking task. They *predicted* the syntactic contexts in which regularized SVA appears in the rating task, as observed in the rating differences between AAL-grammatical items (*We was happy*) and items that violated AAL's SVA rules (*He were happy*). (I have discussed this in terms of *granularity*.) They also *predicted* regularized forms in a neutral context in the transcription task. I did not have tasks that directly corresponded to *labeling* or *using* the form (beyond open-ended questions at the end of the study, including the self-report about general usage of AAL, which did not narrowly target regularized SVA).

As I noted in the introduction to this dissertation, these abilities reflect multiple types of knowledge—morphosyntactic, semantic, and indexical—at multiple levels of granularity. In fact, Labov's approach is probably best viewed as a series of *diagnostics*, rather than a theoretical

framework for mental grammars; all "abilities" are specified in terms of tasks, and these tasks may require different types of knowledge for successful completion. (To be clear, Labov does not claim otherwise; the paper is framed in terms of ways to probe knowledge of other dialects.) Thus, we can now turn to the task of mapping from these diagnostics to questions about mechanisms and representations. While much remains to be learned, the empirical evidence in this dissertation, combined with several decades of research and theorizing in psycholinguistics, equips us to begin this task.

As a starting point, we should consider the source of the knowledge observed throughout this dissertation and other studies. Labov (1973) notes the possibility of multiple avenues to completing the diagnostic tasks he describes:

> Let us consider A', a native speaker of Dialect A, who has been in contact with Dialect B for some time. If other speakers of A have been in intimate contact with B for long periods, [...] the work of translation is apt to be quite simple: a mechanical re-coding of lexical items. But if there is no history of contact, there may be deeper differences, and [the listener] must use his own linguistic competence to ferret them out. He may not know himself how much or how little he knows about this dialect, but we should be able to put him to the test in various ways to decide how much of [a grammatical variant] he has assimilated: whether he has activated his pan-dialectal grammar or constructed one for the occasion. (p. 45)

In principle, there are a few ways that a participant might exhibit apparent knowledge of variation. First, they might be *exposed* to a particular variant (the "contact" referenced by Labov), and build a representation of this variant and the environments in which it may occur. This is essentially what I have implied throughout this dissertation, as I have discussed the idea that participants already know that regularized SVA occurs. The tasks were largely designed in ways that best align with this interpretation; participants *never* received information about SVA rules in each guise, since SVA was either neutral or ambiguous in all audio items. In the rating task, items with regularized SVA were presented among items that are well- or ill-formed in both MAE and AAL, and this task always came last, meaning that participants did not need

to consider the existence of regularized SVA in the transcription or visual world tasks.

However, in some situations, a participant might draw upon their own grammar to complete the task, without any representation of grammars that they do not produce. Although most English speakers are likely exposed to regularized SVA, let us consider the strong case of someone who is not. Perhaps their mental representations of SVA are in some way unstable, and this instability in representations is what gives rise to SVA variation to begin with (Roeper and Green, 2008). For example, perhaps *was* is the default form, while *were* is a marked form, and the bare/zero-marked verb is the default, while verbal *-s* is a marked form.[1] If this is the case, in the special situation of listening to an unfamiliar dialect, a listener might extend their expectations of the default form to a wider range of environments. The present results cannot entirely exclude this explanation, but this seems like a less likely explanation for both the transcription and eye-tracking tasks, since the experimental materials were always ambiguous in ways that could be consistent with the MAE grammar.

However, for other findings, it seems more plausible that participants have used their own grammar to coerce a representation of a novel form. Depending on the available analogous structures, this could lead to diverging accuracy depending on the variable in question. This explains both the success of MAE-speaking participants at interpreting negative auxiliary inversion (Blanchette et al., 2023) and their failure at interpreting stressed *BIN* (Beyer et al., 2015) and habitual *be* (Wolfram, 1982, and Chapter 4 of this dissertation). As Blanchette et al. (2023) argue, participants arrive at the wide scope interpretation of *Didn't nobody like the movie* because it is a "structural analogue" of *Not everybody liked the movie*, which occurs in the same syntactic environments and has the same scope of negation as negative auxiliary inversion. On the other hand, for *BIN* and *be*, the structural analogues are probably *have/has been* and the present

[1] While this is an illustrative hypothetical, it can be independently motivated on a variety of grounds.

tense copula, respectively. These analogues do not have the same aspectual properties, leading to potential differences in truth conditions and poor performance on tasks involving interpretation. Labov and Baker (2015) make a similar argument for young AAL speakers confronted with verbal *-s*; since *-s* is exclusively a plural marker that attaches to nouns in some varieties of AAL, some listeners might attempt to accommodate a verbal *-s* with a plural interpretation, even though it only appears with singular subjects in MAE.

This is potentially orthogonal to the issue of time course, which had a minimal effect throughout the present studies. If these forms are coerced in the moment, we might say that the listener has "constructed [a pan-dialectal grammar] for the occasion," in Labov's terms. In cases like negative auxiliary inversion (Blanchette et al., 2023) and *needs washed* (Kaschak and Glenberg, 2004; Kaschak, 2006), this process seems to occur over the course of the experiment. It is also possible that listeners have already incorporated these forms into their "pan-dialectal" grammar prior to participation in an experiment, but their representation differs from that of speakers who actually use a given variant (e.g., the over-application of habitual *be*).

5.3 Relevance to key questions in psycholinguistics and sociolinguistics

This dissertation has combined methods and frameworks from psycholinguistics and sociolinguistics to better understand the cognition of morphosyntactic variation. While I believe this integrative work is important on its own terms, it can also provide relevant evidence for open areas in each sub-discipline. Since the on-line and off-line results aligned, the results have limited bearing on questions of time course, but they broadly align with constraint-based models (e.g., Spivey-Knowlton and Sedivy, 1995) where top-down social constraints influence initial processing commitments. The findings make the most direct contact with noisy channel models, an area of recent interest in psycholinguistics that makes no predictions regarding time course, and questions

of social evaluation, which have a long history of importance to sociolinguistic research.

5.3.1 Noisy channel models

As I noted in the introduction to this dissertation, noisy channel modeling (e.g., Gibson et al., 2013) has become a common approach in psycholinguistics to explain how listeners arrive at interpretations when there are various sources of noise in the linguistic signal, including factors relating to the speaker, like production errors, and factors relating to the environment, like ambient noise. In this framework, listeners are attempting to infer a speaker's intended meaning from a perceived linguistic form. In some situations, they will infer a meaning that differs from the form they perceived if "noise" is sufficiently probable and the form is sufficiently improbable. To repeat a common example, if a listener perceives (51a), they might infer that the intended meaning was (51b); a deletion of *to* is more likely than the implausible scenario described in the perceived sentence.

(51) a. The mother gave the candle the daughter.

b. The mother gave the candle **to** the daughter.

Under a noisy channel approach, we might think of dialect differences as a source of noise. Unlike most studies in the noisy channel literature, the listener would usually infer that a different *form* was intended, but not a different meaning. For example, if an MAE speaker hears (52a), they might infer that the "intended" form (relative to their own grammar) was (52b), but the truth-conditional meaning is unchanged.

(52) a. The mother **give** the candle to the daughter.

b. The mother **gives** the candle to the daughter.

The transcription task provides a relatively direct test for this idea. If a listener is more

likely to "correct" the differences in a speaker of a different dialect, we might expect counter-intuitive performance among MAE speakers: they should be *more* likely to transcribe MAE SVA in the AAL guise, compared to the MAE guise. In both guises, the splicing procedure meant that the subtle acoustic features favored regularized SVA, meaning a listener might perceive *He sit still*. However, the data from AAL speakers is "noisier," so MAE-dominant listeners should be more likely to override this subtle acoustic bias when listening to an AAL speaker. Instead, I observed the opposite effect, where both listener groups were more likely to transcribe regularized SVA in the AAL guise.

I can see a few ways to hybridize these results with a noisy channel account. First, perhaps the instructions for the task biased participants against a noisy channel inference; they were instructed to transcribe exactly what they heard. On its own, this does not explain the existence of a guise effect in the opposite of the predicted direction, but perhaps it allows noisy channel proponents to discard these data. Second, perhaps noisy channel inferences for differences in form function differently from differences in meaning; listeners assume that a speaker of a different dialect intended a form that differs from that of the listener's dominant dialect. Thus, an MAE-dominant listener would infer that the AAL speaker did not intend to use verbal *-s*, despite having the intended meaning of a singular subject. This could be a promising area for future work to elaborate accounts of noisy channel inferencing.

The noisy channel approach is perhaps better equipped to account for the picture selection data in the eye-tracking task, since this is a situation where meaning and form align. From the perspective of an MAE-dominant listener, the selections of a singular image could reflect the idea that an AAL speaker is more likely to "delete a verbal *-s*" (relative to the MAE grammar), meaning that a listener is more likely to infer that a verbal *-s* was intended, and therefore, a singular interpretation is merited. However, this does not account for the absence of a guise effect

for the noun morphology items.

Thus, across tasks, it is clear that for a noisy channel model to account for these results, a rich "noise model" will be necessary, where MAE-speaking listeners represent the specific ways in which AAL differs from MAE. The results in this dissertation, and in other studies of cross-dialect processing and comprehension, are providing evidence about the content of that noise model. Given the difference in guise effect between the noun morphology and verb morphology items, the primary categories explored thus far—focusing on the likelihood of deletions, insertions, and substitutions (Ryskin et al., 2018)—are inadequate. Even worse, they use an idealized monodialectal grammar as the benchmark for the absence of noise, and this does not reflect the linguistic experience of most people.

This raises the question of how we account for the patterns for bidialectal listeners under a noisy channel account. One possibility is that bidialectal participants are doing something fundamentally different from monodialectal listeners; since all observed forms fall within bidialectal participants' linguistic repertoire, noisy channel models are not relevant. This explanation is questionable, since bidialectal participants' responses were not qualitatively different from those of monodialectals. Instead, I would argue that "noise models" often do not model *noise* at all, but rather a sociolinguistic *signal*. For the step of inferring the intended sentence, a better term would be the *speaker model*, since the work in this dissertation shows that listeners have mental models of *speakers'* grammars. These speaker models are tied to representations of different dialects and can still combine with models of noise in the environment. The present results leave open the question of the degree to which these models are specific to an individual or lect, but the present results suggest that there must be some mapping function between the variants that a listener observes from a given speaker and the listener's pre-existing representations of a lect; this is necessary to explain effects for SVA in guises where no SVA information was directly provided.

5.3.2 Relationship between knowledge of morphosyntactic variation and social evaluation

In addition to broadly suggesting paths for integration between psycholinguistics and sociolinguistics, the present results bear on ongoing questions in sociolinguistics. One area of ongoing research has been social evaluation and the sociolinguistic monitor, where the question is roughly the reverse of what I have pursued here: given some linguistic variant (or rate of usage of a particular variant), what social inferences does a listener draw about a speaker, and what sorts of linguistic variants are most likely to trigger a social evaluation?

The present work expands on findings that morphosyntactic variation is subject to social evaluation (Levon and Buchstaller, 2015), contrary to claims of the Interface Principle, which privileges surface phonetic forms (Labov et al., 1993). The present findings mean that there is now bidirectional evidence that listeners evaluate the social meaning of grammatical variants; they both infer social meanings from usage rates of minoritized grammatical forms (Levon and Buchstaller, 2015) and predict grammatical variants given a broader social/sociolinguistic context (rating and transcription tasks). While my experiments confounded social categories with the presence of other linguistic variants, it is clear from work by Squires (2019), that a simple instruction to participants about context is sufficient to change expectations (in her case, song lyrics).[2]

This does not mean, however, that there is a one-to-one mapping between social categories and linguistic representations. Austen (2020) found that listeners evaluated the prestige variant of the vowel in BATH in the TRAP/BATH split of British English as "posh" — even if it was used in an incorrect context, like GAS, and even if the listener produced the split. This parallels my finding that bidialectal participants still evaluated unattested SVA patterns (*He were happy*) as more likely in AAL than in MAE, which parallels findings by Squires (2014). This could be

[2]My results do not contradict the idea that morphosyntactic and phonetic variables tend to take on different types of social meanings and may be used recruited by speakers in different ways (Eckert, 2019); thus, the Interface Principle might still be a useful first approximation for nuances that were not tested here.

an interesting area for future work on the relationship between the grammatical conditions of a variant and the social inferences drawn from its usage; perhaps unattested forms are more likely to trigger stereotypes associated with AAL than common AAL forms, aligning with findings by Bender (2005) about the AAL copula and findings by Vaughn (2022) about (ING). It is worth considering multiple possible directions for the effect of using an unattested form. When an AAL-dominant speaker uses such a form, it might be viewed as a hypercorrection, just like a production of GAS with the BATH vowel would trigger a social evaluation of "trying to sound posh." On the other hand, when an MAE-dominant speaker uses an unattested form when apparently approximating AAL, the usage of an unattested SVA pattern might more strongly index a social category associated with AAL speakers. The latter case can be understood in terms of linguistic appropriation (e.g., Eberhardt and Freeman, 2015).

With regard to research on the sociolinguistic monitor (Labov et al., 2011), the present results leave open the question of how different usage rates of the regularized variants (and other grammatical variants, like negative concord), are necessary to adjust listeners' expectations. In my studies, participants never heard an unambiguously regularized form in the critical items, but in filler items, they always heard a stereotyped grammatical variant. Thus, I presume that when participants were listening to AAL in my tasks, they had reached an asymptote where all associations with AAL were maximally triggered, as was observed by Labov et al. (2011) once rates of the *-in* variant exceeded around 50%. Thus, increasing the amount of stereotyped AAL features would have had no impact on listeners' social evaluations, but perhaps *decreasing* the proportion of AAL features in the AAL guise could have had an impact.

The present results also bear on questions of lectal coherence in the minds of speakers and listeners (Guy, 2013; Guy and Hinskens, 2016). That is, in Guy and Hinskens's (2016) framing, to what extent do "linguistic variables systematically co-vary"? When speaking of a "dialect"

or "variety," such as AAL or MAE, we assume that the different features of that dialect tend to pattern together, and this was built into the design of my experiments: in filler items, there was a perfect correlation between the usage of negative concord, negative inversion, and null copula, since each guise categorically used either MAE or non-MAE variants. Since participants received no direct information about each voice's SVA rules, the presence of a guise effect means that there is some connection in listeners' minds between the features they heard and regularized SVA. As I discussed above, more work will be necessary to determine the exact categories in listeners' minds for each guise. Throughout this dissertation, I have described the guises in social terms, since they were easily racialized by participants, but the social categories and linguistic categories are undoubtedly interacting in complex ways. We can frame this as the interaction between the syntactic and indexical representations involved in listeners' representations of variation (see Tamminga and Wade, 2022, for a discussion on coherence across different levels of granularity).

On a broader level, sociolinguistics is typically described as the study of "language in its social context." A complete account of this must include the cognitive linguistic resources that speakers bring to multiple social situations, and this dissertation makes it clear that such resources exist and can be studied with a variety of experimental paradigms.

5.4 Future directions

5.4.1 Determining how the present results generalize

This dissertation provides clear evidence that listeners do adjust their morphosyntactic expectation depending on the dialect of the speaker, and that their own linguistic experience affects the ways in which they do so. In the above discussion, I have suggested many areas where subtle design changes would help us to better understand the boundary conditions of the observed effects, and different plausible mechanisms. One simple area of consideration is the

morphosyntactic phenomena under study; I have focused on regularized SVA, but future work might consider other phenomena. It seems that English speakers from the United States have relatively high familiarity with SVA variation, despite a preference for MAE-compatible variants; but this did not apply, for example, to habitual *be*. In selecting other phenomena for study, it could be illuminating to jointly consider listeners' likely level of exposure to a given variable, listeners' level of explicit awareness of forms (i.e. comparing *stereotypes*, *markers*, and *indicators*), and the structural analogues available in the listener's primary dialect. While there has been scattered work across phenomena, it would be helpful for future work to use the same methods and participants to study different phenomena, allowing for stronger inferences about any differences.

Another area where there is considerable space for exploration is the "guise" manipulation. As I noted above, multiple factors were confounded, including the racialization of the speaker and usage of grammatical forms that differ from MAE. Future work could consider a variety of intersections of identity and usage rates of different sociolinguistic variants. Such work should build on the rich literature on social evaluation and consider the potential for different levels of *granularity* in the *indexical* dimension of knowledge. For example, a White monodialectal MAE speaker might have a murky, general representation of AAL, while a Black AAL speaker might have richer representations of variation within AAL according to socioeconomic, regional, generational, and other factors. Additionally, in my experiments, a given voice categorically used a given variant, which does not reflect the usage patterns of many speakers. Future work could consider the impact of different usage rates and the common practice of a given speaker shifting in different contexts.

Moreover, the nature of the experimental context could also be manipulated to modulate participants' expectations regarding the usage of AAL. While there are inherent challenges to creating a felicitous context for AAL in an experimental setting, a decreased reliance on writing

(or usage of AAL spelling conventions) could be helpful, as could the inclusion of multiple AAL speakers in a session; such approaches have yielded interesting results in bilingualism research (Kaan et al., 2020; Litcofsky and Van Hell, 2017). Finally, future work should include a variety of listener backgrounds and ways to characterize listener backgrounds. My studies created an artificial binary between bidialectal and monodialectal listeners, based on a combination of racial identity and self-reported AAL and MAE usage (collapsing from an ordinal scale). This was a deliberate simplification given the limited amount of work in this area, and the observed group effects provide a clear proof-of-concept for further exploration of the sources of variation across participants. This might include data from sociolinguistic interviews, more specific information about participants' social networks, and directly measured or self-reported information about participants' shifting practices.[3] In addition to measure of oral language, literacy measures could also be relevant for understanding participants' experience with different language varieties, particularly in studies that involve written language.[4]

All of these factors—phenomena, speakers, listeners—likely interact in complex ways, and given the paucity of data in this area, much incremental work is necessary before strong claims can be justified about listeners' knowledge and processing of morphosyntactic variation across dialects. Regardless, such work should be done with an eye toward understanding the representations and mechanisms involved.

5.4.2 Larger theoretical questions

As I mentioned above, one question is whether there is any implicational hierarchy for types of knowledge. We know that people with varying degrees of experience with a particular dialect show differing patterns of knowledge for both grammatical conditions and interpretation, and we

[3]Thank you to Jorge Valdés Kroff for the suggestion about switching behavior.
[4]Thank you to Shenika Hankerson for this suggestion.

might expect these different types of knowledge to broadly correlate. But is one type of knowledge a necessary first step? One possibility is that listeners follow a singular trajectory; for example, they might first identify broad patterns of difference, then link this pattern to a social group, then learn the (truth-conditional) meaning of a particular variant, then identify the specifics of its syntactic distribution. A different possibility is that the different components of a variable are sufficiently dissociable, so Person A might know the syntactic distribution but not the meaning, while Person B might know the meaning and not the syntactic distribution. As I noted in Chapter 1, I think that these types of knowledge are logically separable, and group effects were observed for all types of knowledge, so it is difficult to determine any relative rankings.

Another question that this project opens is the switching dynamics between different dialects. I observed general differences in listeners' expectations depending on the grammar of their interlocutor, and there was participant-level variability in the magnitude of guise effects. However, this variability was *not* explained by participants' status as a monodialectal or bidialectal speaker; contrary to my hypotheses, I did not find evidence that bidialectal listeners make more fine-tuned adjustments to the grammar of their interlocutor than do monodialectal participants. Thus, at this point, individual-level variability is unexplained. Future research should consider both "trait" and "state" explanations for differentiated expectations in listening to different dialects (see Salig et al., 2021, for discussion of traits and states in bilingualism research). That is, are there traits that a given listener might have—in terms of linguistic representations or other cognitive resources—that would make them more or less likely to adjust their expectations depending on a given speaker? Additionally, are there some contexts where a given listener might be more inclined to fine-tune their grammatical expectations to a particular dialect, relative to other contexts, where both grammars remain relatively activated? One might imagine, for example, that listeners who maintain a strict boundary between professional and personal settings for each

dialect might fine-tune their expectations, while listeners who are accustomed to code-meshing practices (Young, 2009) might not. For example, the nature of dialect separation practices has been suggested by Vorwerg et al. (2019) to explain differences in lexical switch cost patterns between Swiss bidialectals and Scottish bidialectals (Kirk et al., 2018).

These questions bring work on cognition across dialects into conversation with ongoing developments in the study of bilingualism, where recent work has shown differing neural indices of code-switches depending on the practices of a given speaker's community (e.g., Beatty-Martínez et al., 2021). Ultimately, I envision an approach akin to the *Adaptive Predictability hypothesis* proposed by Valdés Kroff and Dussias (2023) for bilingualism, where listeners "adapt... how they predict upcoming linguistic information, because of accumulated experience with codeswitched speech" and "cognitive control is the primary domain-general cognitive mechanism that supports rapid integration of other-language information in online comprehension" (p. 23). This approach emphasizes the fact that the "listener must consider a trade-off between actively predicting or holding off on predicting upcoming information" (p. 24), and the combination of their experience and the context will dictate how they make this trade-off.

While all of this research is removed from applied work, it also helps to re-frame broader discussions about social justice and dialect variation. Most crucially, it helps to move past "deficit narratives" where the educational discourse about AAL is framed in terms of AAL-speaking children's lack of knowledge of MAE. Instead, knowledge of multiple linguistic forms is viewed as a valuable skill, and more burden can be placed on listeners from dominant groups, such as White monodialectal MAE speakers, in learning to adjust to a dialect that they do not produce. These results also suggest that many apparent "monodialectals" actually have some knowledge of AAL grammatical patterns, even if they might describe AAL as "ungrammatical." This could be informative to future efforts to intervene on teachers' (and others') language ideologies; if most

people already have implicit knowledge that AAL is rule-governed, simply pointing this out is unlikely to yield a change. However, it could serve as a starting point from which an intervention might build.

5.5 Conclusion

Across multiple measures relating to multiple types of knowledge, we have seen that listeners know the rules of regularized subject-verb agreement, associate the phenomenon with AAL, and adjust their grammatical expectations and interpretations according to who is speaking. This is a remarkable level of knowledge, particularly considering that the transcription and eye-tracking tasks involved ambiguity where an MAE-compatible option was always available. Thus, for a given person, both grammatical competence and performance must involve some representation of variation, rather than one unitary idiolect. However, there is also variability between people in their knowledge of morphosyntactic variation. While this can be observed broadly between groups of participants, there is considerable variation within groups, including minoritized groups, such as speakers of AAL. It is important for all future work to consider this in forming theories of how language is represented and processed in the mind, and I hope that this dissertation can serve as a foundation for further research in the cognitive science of language that includes questions of diversity at the center of theory-building.

Appendix A: Materials and models for Experiments 1 and 2

A.1 Experiment 1

A.1.1 Stimuli

A.1.1.1 Audio Clips

Table A.1: AAL guise audio files, Experiment 1 (CORAAL)

Sentence	Key word	Phenomenon	Source
I don't know nobody over there no more	nobody	Negative concord	DCB_se2_ag4_m_01_1
Basically, all I need is a computer, a desk, and a bed	desk	Cluster reduction	VLD_se0_ag4_f_01_1
I had the best of both worlds	both	TH	ATL_se0_ag2_f_01_1
It done got worse	done	Completive done	DCB_se1_ag4_f_01_2
My mother fought for the change	mother	R-less	LES_se0_ag2_f_01_2
I done been everywhere	done	Completive done	ROC_se0_ag1_m_01_1
I never test well	test	Cluster reduction	DCB_se2_ag2_m_01_1
I don't like nothing sweet on my chicken	nothing	Negative concord	DCB_se1_ag2_f_01_1
My brother actually ran away when I was in middle school	brother	R-less	ATL_se0_ag1_m_02_1
You can't just build a building and be like "this is for the youth"	youth	TH	LES_se0_ag3_m_01_2

Table A.2: MAE guise audio files, Experiment 1 (Santa Barbara Corpus)

Sentence	Key word	Phenomenon	Source
I don't see any back in here	any	Single negation	SBC029
But I got so nervous, right when I was taking the test?	test	Full cluster	SBC028

Sentence	Key word	Phenomenon	Source
Amy had her wisdom teeth taken out	teeth	TH	SBC043
You mean you actually thought we had seen this before?	before	Already	SBC028
Her father at that time was the minster of a nearby church	father	R-ful	SBC040
I think I liked her before	before	Already	SBC045
well mine was totally west	west	Full cluster	SBC032
Now you can't do anything to this guy	anything	Single negation	SBC008
I don't know what player it is	player	R-ful	SBC024
I know how we can make money with this	this	TH	SBC017

A.1.1.2 Written Items

A.1.1.3 Verbal *-s*

1.
   - **Grammatical-Both:** They like to play a lot
   - **Grammatical-AAL:** She like to play a lot
   - **Ungrammatical-Agreement:** They likes to play a lot
   - **Ungrammatical-Other:** Them like to play a lot

2.
   - **Grammatical-Both:** They like to see her in the house
   - **Grammatical-AAL:** He like to see her in the house
   - **Ungrammatical-Agreement:** They likes to see her in the house
   - **Ungrammatical-Other:** Him likes to see her in the house

3.
   - **Grammatical-Both:** They look nice, though
   - **Grammatical-AAL:** It look nice, though
   - **Ungrammatical-Agreement:** You looks nice, though
   - **Ungrammatical-Other:** Them look nice, though

4.
   - **Grammatical-Both:** These people look like they don't belong here
   - **Grammatical-AAL:** This person look like they don't belong here
   - **Ungrammatical-Agreement:** These people looks like they don't belong here
   - **Ungrammatical-Other:** Her looks like she don't belong here

5.
   - **Grammatical-Both:** You seem like a big music fan
   - **Grammatical-AAL:** He seem like a big music fan
   - **Ungrammatical-Agreement:** You seems like a big music fan
   - **Ungrammatical-Other:** Him seems like a big music fan

6.
   - **Grammatical-Both:** You seem to be a very wise person
   - **Grammatical-AAL:** She seem to be a very wise person
   - **Ungrammatical-Agreement:** You seems to be a very wise person
   - **Ungrammatical-Other:** Her seems to be a very nice person

7.
   - **Grammatical-Both:** They seem like they should be together
   - **Grammatical-AAL:** It seem like they should be together
   - **Ungrammatical-Agreement:** They seems like they should be together

- **Ungrammatical-Other:** Them seem like they should be together

8. - **Grammatical-Both:** They all want a little water
   - **Grammatical-AAL:** Somebody want a little water
   - **Ungrammatical-Agreement:** They all wants a little water
   - **Ungrammatical-Other:** Them want a little water

9. - **Grammatical-Both:** The people want them to be a mile away
   - **Grammatical-AAL:** The town want them to be a mile away
   - **Ungrammatical-Agreement:** The people wants them to be a mile away
   - **Ungrammatical-Other:** Her wants them to be a mile away

10. - **Grammatical-Both:** We never want to sit around
    - **Grammatical-AAL:** She never want to sit around
    - **Ungrammatical-Agreement:** We never wants to sit around
    - **Ungrammatical-Other:** Her never wants to sit around

11. - **Grammatical-Both:** The cities want that tourist money
    - **Grammatical-AAL:** The city want that tourist money
    - **Ungrammatical-Agreement:** The cities wants that tourist money
    - **Ungrammatical-Other:** Them want that tourist money

12. - **Grammatical-Both:** My kids love comedies
    - **Grammatical-AAL:** My wife love comedies
    - **Ungrammatical-Agreement:** My kids loves comedies
    - **Ungrammatical-Other:** Them love comedies

13. - **Grammatical-Both:** I still love being at school
    - **Grammatical-AAL:** She still love being at school
    - **Ungrammatical-Agreement:** I still loves being at school
    - **Ungrammatical-Other:** Her still loves being at school

14. - **Grammatical-Both:** His neighbors love him
    - **Grammatical-AAL:** His neighborhood love him
    - **Ungrammatical-Agreement:** His neighbors loves him
    - **Ungrammatical-Other:** Them love him

15. - **Grammatical-Both:** My parents feel sorry for me
    - **Grammatical-AAL:** My mother feel sorry for me
    - **Ungrammatical-Agreement:** My parents feels sorry for me
    - **Ungrammatical-Other:** Them feel sorry for me

16. - **Grammatical-Both:** They feel slighted
    - **Grammatical-AAL:** He feel slighted
    - **Ungrammatical-Agreement:** They feels slighted
    - **Ungrammatical-Other:** Him feels slighted

17. - **Grammatical-Both:** They know a lot of songs by heart
    - **Grammatical-AAL:** She know a lot of songs by heart
    - **Ungrammatical-Agreement:** They knows a lot of songs by heart
    - **Ungrammatical-Other:** Them know a lot of songs by heart

18.
- **Grammatical-Both:** They make us come back in the building
- **Grammatical-AAL:** She make us come back in the building
- **Ungrammatical-Agreement:** They makes us come back in the building
- **Ungrammatical-Other:** Her makes us come back in the building

19.
- **Grammatical-Both:** You think you can do that
- **Grammatical-AAL:** He think he can do that
- **Ungrammatical-Agreement:** You thinks you can do that
- **Ungrammatical-Other:** Him thinks he can do that

20.
- **Grammatical-Both:** You know me
- **Grammatical-AAL:** He know me
- **Ungrammatical-Agreement:** You knows me
- **Ungrammatical-Other:** Them know me

21.
- **Grammatical-Both:** They know what to do
- **Grammatical-AAL:** She know what to do
- **Ungrammatical-Agreement:** They knows what to do
- **Ungrammatical-Other:** Her knows what to do

22.
- **Grammatical-Both:** You make good music
- **Grammatical-AAL:** He make good music
- **Ungrammatical-Agreement:** You makes good music
- **Ungrammatical-Other:** Him makes good music

23.
- **Grammatical-Both:** They call you to do something
- **Grammatical-AAL:** She call you to do something
- **Ungrammatical-Agreement:** They calls you to do something
- **Ungrammatical-Other:** Her calls you to do something

24.
- **Grammatical-Both:** They make you embarassed in front of all these people
- **Grammatical-AAL:** He make you embarassed in front of all these people
- **Ungrammatical-Agreement:** They makes you embarassed in front of all these people
- **Ungrammatical-Other:** Him makes you embarassed in front of all these people

25.
- **Grammatical-Both:** The teachers get after us sometimes
- **Grammatical-AAL:** The teacher get after us sometimes
- **Ungrammatical-Agreement:** The teachers gets after us sometimes
- **Ungrammatical-Other:** Them get after us sometimes

26.
- **Grammatical-Both:** They come in my house sometimes
- **Grammatical-AAL:** He come in my house sometimes
- **Ungrammatical-Agreement:** They comes in my house sometimes
- **Ungrammatical-Other:** Them come in my house sometimes

27.
- **Grammatical-Both:** You talk all funny
- **Grammatical-AAL:** He talk all funny
- **Ungrammatical-Agreement:** You talks all funny
- **Ungrammatical-Other:** Him talks all funny

28.
- **Grammatical-Both:** They never get paid
- **Grammatical-AAL:** He never get paid
- **Ungrammatical-Agreement:** They never gets paid
- **Ungrammatical-Other:** Them never get paid

29.
- **Grammatical-Both:** You take your time
- **Grammatical-AAL:** She take her time
- **Ungrammatical-Agreement:** You takes your time
- **Ungrammatical-Other:** Her takes her time

30.
- **Grammatical-Both:** They get what you mean
- **Grammatical-AAL:** He get what you mean
- **Ungrammatical-Agreement:** They gets what you mean
- **Ungrammatical-Other:** Him gets what you mean

31.
- **Grammatical-Both:** We type something on the computer
- **Grammatical-AAL:** He type something on the computer
- **Ungrammatical-Agreement:** We types something on the computer
- **Ungrammatical-Other:** Us types something on the computer

32.
- **Grammatical-Both:** We smoke cigarettes after work
- **Grammatical-AAL:** He smoke cigarettes after work
- **Ungrammatical-Agreement:** We smokes cigarettes after work
- **Ungrammatical-Other:** Us smokes cigarettes after work

A.1.1.4 (WAS)

1.
- **Grammatical-Both:** He was tearing the house down
- **Grammatical-AAL:** They was tearing the house down
- **Ungrammatical-Agreement:** He were tearing the house down
- **Ungrammatical-Other:** He was tear the house down

2.
- **Grammatical-Both:** She was just looking at me
- **Grammatical-AAL:** They was just looking at me
- **Ungrammatical-Agreement:** She were just looking at me
- **Ungrammatical-Other:** She was just look at me

3.
- **Grammatical-Both:** This boy was breaking the machine
- **Grammatical-AAL:** These boys was breaking the machine
- **Ungrammatical-Agreement:** This boy were breaking the machine
- **Ungrammatical-Other:** This boy was break the machine

4.
- **Grammatical-Both:** The kid was going to school
- **Grammatical-AAL:** The kids was going to school
- **Ungrammatical-Agreement:** The kid were going to school
- **Ungrammatical-Other:** The kid was go to school

5.
- **Grammatical-Both:** The doctor was rushing in
- **Grammatical-AAL:** The doctors was rushing in
- **Ungrammatical-Agreement:** The doctor were rushing in
- **Ungrammatical-Other:** The doctor was rush in

6. • **Grammatical-Both:** She was putting on a lot of plays
   • **Grammatical-AAL:** They was putting on a lot of plays
   • **Ungrammatical-Agreement:** She were putting on a lot of plays
   • **Ungrammatical-Other:** She was put on a lot of plays

7. • **Grammatical-Both:** My brother was heading out the door
   • **Grammatical-AAL:** My brothers was heading out the door
   • **Ungrammatical-Agreement:** My brother were heading out the door
   • **Ungrammatical-Other:** My brother was head out the door

8. • **Grammatical-Both:** He was waiting on the other side
   • **Grammatical-AAL:** They was waiting on the other side
   • **Ungrammatical-Agreement:** He were waiting on the other side
   • **Ungrammatical-Other:** He was wait on the other side

9. • **Grammatical-Both:** She was just taking the money
   • **Grammatical-AAL:** They was just taking the money
   • **Ungrammatical-Agreement:** She were just taking the money
   • **Ungrammatical-Other:** She was just take the money

10. • **Grammatical-Both:** I was crying
    • **Grammatical-AAL:** They was crying
    • **Ungrammatical-Agreement:** I were crying
    • **Ungrammatical-Other:** I was cry

11. • **Grammatical-Both:** She was having trouble
    • **Grammatical-AAL:** They was having trouble
    • **Ungrammatical-Agreement:** She were having trouble
    • **Ungrammatical-Other:** She was have trouble

12. • **Grammatical-Both:** I was singing well
    • **Grammatical-AAL:** They was singing well
    • **Ungrammatical-Agreement:** I were singing well
    • **Ungrammatical-Other:** I was sing well

13. • **Grammatical-Both:** He was riding to the grave site
    • **Grammatical-AAL:** They was riding to the grave site
    • **Ungrammatical-Agreement:** He were riding to the grave site
    • **Ungrammatical-Other:** He was ride to the grave site

14. • **Grammatical-Both:** She was stopping to help him
    • **Grammatical-AAL:** We was stopping to help him
    • **Ungrammatical-Agreement:** She were stopping to help him
    • **Ungrammatical-Other:** She was stop to help him

15. • **Grammatical-Both:** He was just trying to adjust
    • **Grammatical-AAL:** We was just trying to adjust
    • **Ungrammatical-Agreement:** He were just trying to adjust
    • **Ungrammatical-Other:** He was just try to adjust

16. • **Grammatical-Both:** I was going out of town

- **Grammatical-AAL:** We was going out of town
- **Ungrammatical-Agreement:** I were going out of town
- **Ungrammatical-Other:** I was go out of town

17.
- **Grammatical-Both:** She was going to a picnic
- **Grammatical-AAL:** They was going to a picnic
- **Ungrammatical-Agreement:** She were going to a picnic
- **Ungrammatical-Other:** She was go to a picnic

18.
- **Grammatical-Both:** The girl was dancing
- **Grammatical-AAL:** These girls was dancing
- **Ungrammatical-Agreement:** The girl were dancing
- **Ungrammatical-Other:** The girl was dance

19.
- **Grammatical-Both:** He was watching us
- **Grammatical-AAL:** They was watching us
- **Ungrammatical-Agreement:** He were watching us
- **Ungrammatical-Other:** He was watch us

20.
- **Grammatical-Both:** He was selling stuff on the corner
- **Grammatical-AAL:** They was selling stuff on the corner
- **Ungrammatical-Agreement:** He were selling stuff on the corner
- **Ungrammatical-Other:** He was sell stuff on the corner

21.
- **Grammatical-Both:** I was wearing bracelets
- **Grammatical-AAL:** The girls was wearing bracelets
- **Ungrammatical-Agreement:** I were wearing bracelets
- **Ungrammatical-Other:** I was wearing bracelets

22.
- **Grammatical-Both:** The team was playing in the playoffs
- **Grammatical-AAL:** They was playing in the playoffs
- **Ungrammatical-Agreement:** The team were playing in the playoffs
- **Ungrammatical-Other:** The team was play in the playoffs

23.
- **Grammatical-Both:** I was riding with him
- **Grammatical-AAL:** They was riding with him
- **Ungrammatical-Agreement:** I were riding with him
- **Ungrammatical-Other:** I was ride with him

24.
- **Grammatical-Both:** She was trying to get us to study
- **Grammatical-AAL:** They was trying to get us to study
- **Ungrammatical-Agreement:** She were trying to get us to study
- **Ungrammatical-Other:** She was try to get us to study

25.
- **Grammatical-Both:** He always was fighting at night
- **Grammatical-AAL:** They always was fighting at night
- **Ungrammatical-Agreement:** He always were fighting at night
- **Ungrammatical-Other:** He always was fight at night

26.
- **Grammatical-Both:** I felt that she was leaning towards that
- **Grammatical-AAL:** I felt that you was leaning towards that

- **Ungrammatical-Agreement:** I felt that she were leaning towards that
- **Ungrammatical-Other:** I felt that she was lean towards that

27. - **Grammatical-Both:** I heard she was messing with my boyfriend
    - **Grammatical-AAL:** I heard you was messing with my boyfriend
    - **Ungrammatical-Agreement:** I heard she were messing with my boyfriend
    - **Ungrammatical-Other:** I heard she was mess with my boyfriend

28. - **Grammatical-Both:** He was lugging that bag
    - **Grammatical-AAL:** You was lugging that bag
    - **Ungrammatical-Agreement:** He were lugging that bag
    - **Ungrammatical-Other:** He was lug that bag

29. - **Grammatical-Both:** He was slipping
    - **Grammatical-AAL:** You was slipping
    - **Ungrammatical-Agreement:** He were slipping
    - **Ungrammatical-Other:** He was slip

30. - **Grammatical-Both:** She was letting me know
    - **Grammatical-AAL:** You was letting me know
    - **Ungrammatical-Agreement:** She were letting me know
    - **Ungrammatical-Other:** She was let me know

31. - **Grammatical-Both:** My friend was going to private school
    - **Grammatical-AAL:** My friends was going to private school
    - **Ungrammatical-Agreement:** My friend were going to private school
    - **Ungrammatical-Other:** My friend was go to private school

32. - **Grammatical-Both:** I was just speaking
    - **Grammatical-AAL:** You was just speaking
    - **Ungrammatical-Agreement:** I were just speaking
    - **Ungrammatical-Other:** I was just speak

A.1.1.5 Null copula

1. - **Grammatical-Both:** I'm clapping for you
   - **Grammatical-AAL:** They clapping for you
   - **Marginal-AAL:** The crowd clapping for you
   - **Ungrammatical-Agreement:** I clapping for you
   - **Ungrammatical-Other:** They'll clapping for you

2. - **Grammatical-Both:** I'm coming for her
   - **Grammatical-AAL:** They coming for her
   - **Marginal-AAL:** Those girls coming for her
   - **Ungrammatical-Agreement:** I coming for her
   - **Ungrammatical-Other:** They'll coming for her

3. - **Grammatical-Both:** I'm enjoying the city
   - **Grammatical-AAL:** They enjoying the city
   - **Marginal-AAL:** That family enjoying the city
   - **Ungrammatical-Agreement:** I enjoying the city

- **Ungrammatical-Other:** They'll enjoying the city

4.
- **Grammatical-Both:** I'm paying them back
- **Grammatical-AAL:** They paying me back
- **Marginal-AAL:** My friend paying me back
- **Ungrammatical-Agreement:** I paying them back
- **Ungrammatical-Other:** They'll paying me back

5.
- **Grammatical-Both:** They won't know what I'm saying
- **Grammatical-AAL:** We won't know what they saying
- **Marginal-AAL:** We won't know what the people saying
- **Ungrammatical-Agreement:** They won't know what I saying
- **Ungrammatical-Other:** We won't know what they'll saying

6.
- **Grammatical-Both:** I'm doing the right thing
- **Grammatical-AAL:** They doing the right thing
- **Marginal-AAL:** My son doing the right thing
- **Ungrammatical-Agreement:** I doing the right thing
- **Ungrammatical-Other:** They'll doing the right thing

7.
- **Grammatical-Both:** That's what I'm doing around here
- **Grammatical-AAL:** That's what they doing around here
- **Marginal-AAL:** That's what the kids doing around here
- **Ungrammatical-Agreement:** That's what I doing around here
- **Ungrammatical-Other:** That's what they'll doing around here

8.
- **Grammatical-Both:** I'm knocking down all of the walls
- **Grammatical-AAL:** They knocking down all of the walls
- **Marginal-AAL:** The crew knocking down all of the walls
- **Ungrammatical-Agreement:** I knocking down all of the walls
- **Ungrammatical-Other:** They'll knocking down all of the walls

9.
- **Grammatical-Both:** I'm laughing
- **Grammatical-AAL:** They laughing
- **Marginal-AAL:** The group laughing
- **Ungrammatical-Agreement:** I laughing
- **Ungrammatical-Other:** They'll laughing

10.
- **Grammatical-Both:** I'm doing things I shouldn't be doing
- **Grammatical-AAL:** They doing things they shouldn't be doing
- **Marginal-AAL:** My kids doing things they shouldn't be doing
- **Ungrammatical-Agreement:** I doing things I shouldn't be doing
- **Ungrammatical-Other:** They'll doing things they shouldn't be doing

11.
- **Grammatical-Both:** I'm trying to open up shop here
- **Grammatical-AAL:** They trying to open up shop here
- **Marginal-AAL:** The barber trying to open up shop here
- **Ungrammatical-Agreement:** I trying to open up shop here
- **Ungrammatical-Other:** They'll trying to open up shop here

12.
- **Grammatical-Both:** They sometimes don't understand what I'm saying

- **Grammatical-AAL:** I sometimes don't understand what they saying
- **Marginal-AAL:** I sometimes don't understand what the bus driver saying
- **Ungrammatical-Agreement:** They sometimes don't understand what I saying
- **Ungrammatical-Other:** I sometimes don't understand what they'll saying

13.
- **Grammatical-Both:** They think I'm going to the truck stop out here
- **Grammatical-AAL:** I think they going to the truck stop out here
- **Marginal-AAL:** I think the drivers going to the truck stop out here
- **Ungrammatical-Agreement:** They think I going to the truck stop out here
- **Ungrammatical-Other:** I think they'll going to the truck stop out here

14.
- **Grammatical-Both:** I'm working during the day
- **Grammatical-AAL:** He working during the day
- **Marginal-AAL:** My dad working during the day
- **Ungrammatical-Agreement:** I working during the day
- **Ungrammatical-Other:** He'll working during the day

15.
- **Grammatical-Both:** I'm coming for your friend
- **Grammatical-AAL:** He coming for your friend
- **Marginal-AAL:** The bully coming for your friend
- **Ungrammatical-Agreement:** I coming for your friend
- **Ungrammatical-Other:** He'll coming for your friend

16.
- **Grammatical-Both:** I'm covering it all up
- **Grammatical-AAL:** She covering it all up
- **Marginal-AAL:** The painter covering it all up
- **Ungrammatical-Agreement:** I covering it all up
- **Ungrammatical-Other:** She'll covering it all up

17.
- **Grammatical-Both:** I'm helping her out
- **Grammatical-AAL:** She helping me out
- **Marginal-AAL:** My uncle helping me out
- **Ungrammatical-Agreement:** I helping her out
- **Ungrammatical-Other:** She'll helping me out

18.
- **Grammatical-Both:** I'm trying to become a nurse
- **Grammatical-AAL:** He trying to become a nurse
- **Marginal-AAL:** The student trying to become a nurse
- **Ungrammatical-Agreement:** I trying to become a nurse
- **Ungrammatical-Other:** He'll trying to become a nurse

19.
- **Grammatical-Both:** I'm living right there
- **Grammatical-AAL:** She living right there
- **Marginal-AAL:** Her grandma living right there
- **Ungrammatical-Agreement:** I living right there
- **Ungrammatical-Other:** She'll living right there

20.
- **Grammatical-Both:** I'm doing too much
- **Grammatical-AAL:** She doing too much
- **Marginal-AAL:** My mom doing too much
- **Ungrammatical-Agreement:** I doing too much

- **Ungrammatical-Other:** She'll doing too much

21.
- **Grammatical-Both:** I'm going to fifth grade
- **Grammatical-AAL:** She going to fifth grade
- **Marginal-AAL:** The girl going to fifth grade
- **Ungrammatical-Agreement:** I going to fifth grade
- **Ungrammatical-Other:** She'll going to fifth grade

22.
- **Grammatical-Both:** I'm getting myself ready to meet her
- **Grammatical-AAL:** She getting herself ready to meet me
- **Marginal-AAL:** My date getting herself ready to meet me
- **Ungrammatical-Agreement:** I getting myself ready to meet her
- **Ungrammatical-Other:** She'll getting herself ready to meet me

23.
- **Grammatical-Both:** I'm making you eat healthy
- **Grammatical-AAL:** She making us eat healthy
- **Marginal-AAL:** My wife making us eat healthy
- **Ungrammatical-Agreement:** I making you eat healthy
- **Ungrammatical-Other:** She'll making us eat healthy

24.
- **Grammatical-Both:** I'm just listening
- **Grammatical-AAL:** He just listening
- **Marginal-AAL:** The pastor just listening
- **Ungrammatical-Agreement:** I just listening
- **Ungrammatical-Other:** He'll just listening

25.
- **Grammatical-Both:** I'm walking around my old neighborhood
- **Grammatical-AAL:** He walking around his old neighborhood
- **Marginal-AAL:** The man walking around his old neighborhood
- **Ungrammatical-Agreement:** I walking around my old neighborhood
- **Ungrammatical-Other:** He'll walking around his old neighborhood

26.
- **Grammatical-Both:** I'm telling the truth
- **Grammatical-AAL:** He telling the truth
- **Marginal-AAL:** His brother telling the truth
- **Ungrammatical-Agreement:** I telling the truth
- **Ungrammatical-Other:** He'll telling the truth

27.
- **Grammatical-Both:** I'm eating a nice meal
- **Grammatical-AAL:** They eating a nice meal
- **Marginal-AAL:** The family eating a nice meal
- **Ungrammatical-Agreement:** I eating a nice meal
- **Ungrammatical-Other:** They'll eating a nice meal

28.
- **Grammatical-Both:** I'm looking for a certain person
- **Grammatical-AAL:** They looking for a certain person
- **Marginal-AAL:** Her sisters looking for a certain person
- **Ungrammatical-Agreement:** I looking for a certain person
- **Ungrammatical-Other:** They'll looking for a certain person

29.
- **Grammatical-Both:** I'm digesting the food

- **Grammatical-AAL:** He digesting the food
- **Marginal-AAL:** The kids digesting the food
- **Ungrammatical-Agreement:** I digesting the food
- **Ungrammatical-Other:** He'll digesting the food

30.
- **Grammatical-Both:** I'm working with her now
- **Grammatical-AAL:** He working with her now
- **Marginal-AAL:** The team working with her now
- **Ungrammatical-Agreement:** I working with her now
- **Ungrammatical-Other:** He'll working with her now

31.
- **Grammatical-Both:** I'm yelling at the person walking past
- **Grammatical-AAL:** He yelling at the person walking past
- **Marginal-AAL:** His boss yelling at the person walking past
- **Ungrammatical-Agreement:** I yelling at the person walking past
- **Ungrammatical-Other:** He'll yelling at the person walking past

32.
- **Grammatical-Both:** I'm riding in a car
- **Grammatical-AAL:** He riding in a car
- **Marginal-AAL:** That man riding in a car
- **Ungrammatical-Agreement:** I riding in a car
- **Ungrammatical-Other:** He'll riding in a car

33.
- **Grammatical-Both:** I'm working during the day
- **Grammatical-AAL:** He working during the day
- **Marginal-AAL:** His daughter working during the day
- **Ungrammatical-Agreement:** I working during the day
- **Ungrammatical-Other:** He'll working during the day

34.
- **Grammatical-Both:** I'm trying to make it
- **Grammatical-AAL:** They trying to make it
- **Marginal-AAL:** Her cousins trying to make it
- **Ungrammatical-Agreement:** I trying to make it
- **Ungrammatical-Other:** They'll trying to make it

35.
- **Grammatical-Both:** I'm always fussing
- **Grammatical-AAL:** She always fussing
- **Marginal-AAL:** My mom always fussing
- **Ungrammatical-Agreement:** I always fussing
- **Ungrammatical-Other:** She'll always fussing

36.
- **Grammatical-Both:** I'm going crazy right now
- **Grammatical-AAL:** He going crazy right now
- **Marginal-AAL:** The teacher going crazy right now
- **Ungrammatical-Agreement:** I going crazy right now
- **Ungrammatical-Other:** He'll going crazy right now

37.
- **Grammatical-Both:** I'm meeting my four friends
- **Grammatical-AAL:** He meeting his four friends
- **Marginal-AAL:** My brother meeting his four friends
- **Ungrammatical-Agreement:** I meeting my four friends

- **Ungrammatical-Other:** He'll meeting his four friends

38.
- **Grammatical-Both:** I'm chilling right now
- **Grammatical-AAL:** He chilling right now
- **Marginal-AAL:** The artist chilling right now
- **Ungrammatical-Agreement:** I chilling right now
- **Ungrammatical-Other:** He'll chilling right now

39.
- **Grammatical-Both:** I'm taking so much money
- **Grammatical-AAL:** They taking so much money
- **Marginal-AAL:** The government taking so much money
- **Ungrammatical-Agreement:** I taking so much money
- **Ungrammatical-Other:** They'll taking so much money

40.
- **Grammatical-Both:** I'm hiding
- **Grammatical-AAL:** She hiding
- **Marginal-AAL:** That person hiding
- **Ungrammatical-Agreement:** I hiding
- **Ungrammatical-Other:** She'll hiding

A.1.2 Model

Table A.3: Regression table for Experiment 1

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	-0.28	0.19	0.01	21.19	-1.50	0.149	
Cond_eff	-1.10	0.23	-0.27	23.95	-4.71	<0.001	***
Guise_eff	-1.80	0.32	-0.43	21.12	-5.65	<0.001	***
Cond_eff:Guise_eff	1.06	0.19	0.13	115.17	5.52	<0.001	***

A.2 Experiment 2

A.2.1 Stimuli

A.2.1.1 Verbal *-s*

1. He commit some crimes
2. He create space for that
3. The girl create something pretty.
4. The child get so mad
5. He get scared all the time
6. My mom keep stuff up there.
7. The dog keep sitting right here
8. He speak Spanish
9. He take so long to get ready
10. She take some time
11. She speak several languages

12. He break some laws
13. She break some of the things away
14. She comfort some families
15. He sit sometimes in the park
16. He sit still for hours
17. She visit sometimes
18. She seek some new opportunities there

A.2.1.2 (WAS)

1. Last week, they was definitely fighting
2. The roads was too slick last night.
3. Things was pretty hot during the war
4. Things was cheaper back then.
5. At that point, my friends was from the area
6. On Sunday, we was back in church.
7. Our lives was great in those days.
8. My parents was getting older then.
9. Years ago, the kids was pretty close.
10. Those relationships was just really close in those days.
11. We was just so taken aback.
12. The neighborhoods was close at one point.
13. They was funding the program for a while.
14. You was twenty at the time
15. They was getting ready to build it that year
16. They was born in the city
17. Back in school, my other friends was in band.
18. I liked it when you was there

A.2.2 Fillers

A.2.2.1 Null Copula

Note: These items were only used for Experiments 2A and 2B.

1. He('s) towing the boat
2. She('s) putting away the broom
3. He('s) cooking chicken
4. She('s) watching the clock
5. He('s) trying the guitar
6. He('s) wearing a helmet
7. She('s) petting the horse
8. He('s) chopping with the knife
9. He('s) catching the mouse
10. She('s) looking at the phone
11. He('s) avoiding the sun
12. She('s) flushing the toilet

A.2.2.2 Negative Inversion

Items used in all experiments:

1. • **AAL:** Can't nobody stop you from trying
   • **MAE:** Nobody can stop you from trying
2. • **AAL:** Can't nobody afford that
   • **MAE:** Nobody can afford that
3. • **AAL:** Can't nobody make nobody do nothing
   • **MAE:** Nobody can make anyone do anything
4. • **AAL:** Don't nobody wanna hear what you gotta say
   • **MAE:** Nobody would hear what you have to say
5. • **AAL:** Don't nobody even know that this is exercise
   • **MAE:** Nobody would know that this is exercise
6. • **AAL:** Ain't nobody gonna bother you there
   • **MAE:** Nobody will bother you here
7. • **AAL:** Ain't nobody else doing it
   • **MAE:** Nobody else will do it
8. • **AAL:** Ain't nobody gonna survive that long
   • **MAE:** Nobody would survive that long
9. • **AAL:** Don't nobody know that feeling
   • **MAE:** Nobody would know that feeling
10. • **AAL:** Ain't nobody knew it
    • **MAE:** Nobody knew it
11. • **AAL:** Ain't nobody getting paid
    • **MAE:** Nobody was getting paid
12. • **AAL:** Ain't nobody here
    • **MAE:** Nobody is here

Items added for Experiment 2C:

1. • **AAL:** Don't nobody really wanna volunteer no more
   • **MAE:** Nobody wanted to volunteer anymore
2. • **AAL:** Can't nobody take that away from you
   • **MAE:** Nobody can take that away from you
3. • **AAL:** Ain't nobody gonna bother you there
   • **MAE:** Nobody will bother you there
4. • **AAL:** Didn't nobody believe I was hurt
   • **MAE:** Nobody believed I was hurt

5. • **AAL:** Don't nobody want to be from there
   • **MAE:** Nobody would want to be from there

6. • **AAL:** Ain't nobody else on that team gonna say that
   • **MAE:** Nobody else on that team would say that

7. • **AAL:** Ain't nobody about to change that
   • **MAE:** Nobody will change that

8. • **AAL:** Couldn't nobody go through that town
   • **MAE:** Nobody could go through that town

A.2.2.3 Negative Concord

Items used in all experiments:

1. He didn't read {no/any} book
2. He didn't skim {no/any} book
3. They didn't bite {no/any} cookie
4. They shouldn't split {no/any} cookie
5. They don't want {no/any} garden
6. He didn't grow {no/any} garden
7. He won't write {no/any} letter
8. They won't sign {no/any} letter
9. That didn't cost {no/any} money
10. They didn't spend {no/any} money
11. She didn't have {no/any} pizza
12. They don't serve {no/any} pizza

Items added for Experiment 2C:

1. They don't hear {no/any} music
2. They can't make {no/any} music
3. He couldn't taste {no/any} sugar
4. They shouldn't add {no/any} sugar

A.2.3 Models

A.2.3.1 Experiment 2A

Table A.4: Regression table for Experiment 2A, order not included

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-2.26	0.32	-2.41	-7.11	<0.001	***
Guise	-3.23	0.52	-1.63	-6.26	<0.001	***
ItemType	3.00	0.36	1.51	8.27	<0.001	***
Guise:ItemType	1.03	0.67	0.26	1.54	0.125	

Table A.5: Regression table for Experiment 2A, order included (effects coded)

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-2.25	0.31	-2.39	-7.23	<0.001	***
Guise	-3.15	0.50	-1.57	-6.24	<0.001	***
ItemType	2.89	0.37	1.45	7.84	<0.001	***
FirstGuise	-0.94	0.48	-0.47	-1.94	0.053	.
Guise:ItemType	0.70	0.68	0.19	1.03	0.304	
Guise:FirstGuise	-1.26	0.68	-0.29	-1.85	0.065	.
ItemType:FirstGuise	-0.31	0.41	-0.09	-0.76	0.446	
Guise:ItemType:FirstGuise	-2.73	0.81	-0.34	-3.38	<0.001	***

Table A.6: Regression table for Experiment 2A, order included (dummy coded)

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-1.53	0.37	-1.53	-4.16	<0.001	***
GuiseMAE	-3.55	0.75	-3.55	-4.71	<0.001	***
ItemTypewas	2.01	0.39	2.01	5.19	<0.001	***
FirstGuiseMAE	-0.83	0.46	-0.83	-1.79	0.074	.
GuiseMAE:ItemTypewas	2.07	0.76	2.07	2.72	0.007	**
GuiseMAE:FirstGuiseMAE	0.10	0.87	0.10	0.11	0.91	
ItemTypewas:FirstGuiseMAE	1.05	0.40	1.05	2.66	0.008	**
GuiseMAE:ItemTypewas:FirstGuiseMAE	-2.73	0.81	-2.73	-3.38	<0.001	***

A.2.3.2 Experiment 2B

Table A.7: Regression table for Experiment 2B

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-2.17	0.26	-2.28	-8.30	<0.001	***
Guise	-2.16	0.41	-1.08	-5.28	<0.001	***
ItemType	2.61	0.41	1.31	6.30	<0.001	***
Guise:ItemType	0.12	0.68	0.03	0.18	0.858	

A.2.3.3 Experiment 2C

Table A.8: Regression table for Experiment 2C, including effect of phenomenon

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-2.92	0.38	-2.94	-7.63	<0.001	***

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
Guise	-2.12	0.40	-1.05	-5.35	<0.001	***
ItemType	0.35	0.51	0.17	0.69	0.487	
Guise:ItemType	-1.10	0.75	-0.28	-1.47	0.142	

Table A.9: Regression table for Experiment 2C, including effect of gender

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-2.96	0.39	-2.98	-7.51	<0.001	***
Guise	-2.17	0.40	-1.09	-5.42	<0.001	***
GuiseGender	-0.29	0.38	-0.14	-0.76	0.447	
Guise:GuiseGender	0.06	0.75	0.02	0.08	0.934	

Appendix B: Materials and models for Experiment 3

B.1 Rating Task

B.1.1 Materials

B.1.1.1 Audio Clips

Table B.1: AAL guise audio files in the sentence rating task

Sentence	Key word	Phenomenon	Source
She ain't buy no jacket	ain't	Negative concord	Transcription task speaker
Basically, all I need is a computer, a desk, and a bed	desk	Cluster reduction	VLD_se0_ag4_f_01_1
Ain't nobody gon' pick them	nobody	Negative inversion	Transcription task speaker
I had the best of both worlds	both	TH	ATL_se0_ag2_f_01_1
It done got worse	done	Completive done	DCB_se1_ag4_f_01_2
My mother fought for the change	mother	R-less	LES_se0_ag2_f_01_2
I done been everywhere	done	Completive done	ROC_se0_ag1_m_01_1
I never test well	test	Cluster reduction	DCB_se2_ag2_m_01_1
I don't like nothing sweet on my chicken	nothing	Negative concord	DCB_se1_ag2_f_01_1
My brother actually ran away when I was in middle school	brother	R-less	ATL_se0_ag1_m_02_1
You can't just build a building and be like "this is for the youth"	youth	TH	LES_se0_ag3_m_01_2

Table B.2: MAE guise audio files in the sentence rating task

Sentence	Key word	Phenomenon	Source
He didn't pour any coffee	any	Single negation	Transcription task speaker
But I got so nervous, right when I was taking the test?	test	Full cluster	SBC028
Nobody knew me	nobody	Single negation, quantifier subject	Transcription task speaker
Amy had her wisdom teeth taken out	teeth	TH	SBC043
You mean you actually thought we had seen this before?	before	Already	SBC028
Her father at that time was the minster of a nearby church	father	R-ful	SBC040
I think I liked her before	before	Already	SBC045
well mine was totally west	west	Full cluster	SBC032
Now you can't do anything to this guy	anything	Single negation	SBC008
I don't know what player it is	player	R-ful	SBC024
I know how we can make money with this	this	TH	SBC017

B.1.1.2 Written Items

Same as items as Experiment 1

B.1.2 Models

B.1.2.1 Bidialectal Participants

Table B.3: Regression table for bidialectal participants in the rating task

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	0.58	0.15	0.32	26.77	3.75	<0.001	***
Cond_forward1	-0.38	0.11	-0.18	2393.78	-3.39	<0.001	***
Cond_forward2	2.02	0.11	0.92	2425.03	17.78	<0.001	***
Cond_forward3	0.41	0.11	0.19	2426.93	3.62	<0.001	***
GuiseMAE	-1.38	0.18	-0.63	30.29	-7.58	<0.001	***
Cond_forward1:GuiseMAE	3.19	0.16	1.45	2403.14	19.77	<0.001	***
Cond_forward2:GuiseMAE	-1.56	0.16	-0.71	2461.47	-9.63	<0.001	***
Cond_forward3:GuiseMAE	-0.26	0.16	-0.12	2457.12	-1.58	0.114	

Table B.4: Post hoc pairwise comparisons for bidialectal participants by grammaticality condition and guise in the rating task

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
Grammatical-Both AAL	Grammatical-AAL AAL	-0.38	0.11	2393.78	-3.39	0.02
Grammatical-Both AAL	Ungrammatical-Agreement AAL	1.64	0.11	2411.92	14.43	0.001
Grammatical-Both AAL	Ungrammatical-Other AAL	2.05	0.11	2409.77	18.06	0.001
Grammatical-Both AAL	Grammatical-Both MAE	-0.16	0.21	50.65	-0.79	0.99
Grammatical-Both AAL	Grammatical-AAL MAE	2.64	0.21	50.65	12.72	0.001
Grammatical-Both AAL	Ungrammatical-Agreement MAE	3.10	0.21	50.65	14.96	0.001
Grammatical-Both AAL	Ungrammatical-Other MAE	3.26	0.21	50.80	15.70	0.001
Grammatical-AAL AAL	Ungrammatical-Agreement AAL	2.02	0.11	2425.03	17.78	0.001
Grammatical-AAL AAL	Ungrammatical-Other AAL	2.43	0.11	2427.49	21.41	0.001
Grammatical-AAL AAL	Grammatical-Both MAE	0.22	0.21	50.69	1.06	0.96
Grammatical-AAL AAL	Grammatical-AAL MAE	3.02	0.21	50.68	14.57	0.001
Grammatical-AAL AAL	Ungrammatical-Agreement MAE	3.49	0.21	50.72	16.81	0.001
Grammatical-AAL AAL	Ungrammatical-Other MAE	3.64	0.21	50.83	17.55	0.001
Ungrammatical-Agreement AAL	Ungrammatical-Other AAL	0.41	0.11	2426.93	3.62	0.007
Ungrammatical-Agreement AAL	Grammatical-Both MAE	-1.80	0.21	50.63	-8.69	0.001
Ungrammatical-Agreement AAL	Grammatical-AAL MAE	1.00	0.21	50.64	4.82	0.001
Ungrammatical-Agreement AAL	Ungrammatical-Agreement MAE	1.46	0.21	50.73	7.06	0.001
Ungrammatical-Agreement AAL	Ungrammatical-Other MAE	1.62	0.21	50.84	7.80	0.001
Ungrammatical-Other AAL	Grammatical-Both MAE	-2.21	0.21	50.69	-10.67	0.001
Ungrammatical-Other AAL	Grammatical-AAL MAE	0.59	0.21	50.62	2.83	0.11
Ungrammatical-Other AAL	Ungrammatical-Agreement MAE	1.05	0.21	50.69	5.08	0.001

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
Ungrammatical-Other AAL	Ungrammatical-Other MAE	1.21	0.21	50.91	5.82	0.001
Grammatical-Both MAE	Grammatical-AAL MAE	2.80	0.11	2390.84	24.49	0.001
Grammatical-Both MAE	Ungrammatical-Agreement MAE	3.27	0.11	2416.26	28.41	0.001
Grammatical-Both MAE	Ungrammatical-Other MAE	3.42	0.11	2404.16	29.86	0.001
Grammatical-AAL MAE	Ungrammatical-Agreement MAE	0.47	0.11	2410.32	4.06	0.001
Grammatical-AAL MAE	Ungrammatical-Other MAE	0.62	0.12	2419.00	5.38	0.001
Ungrammatical-Agreement MAE	Ungrammatical-Other MAE	0.16	0.12	2416.28	1.35	0.88

B.1.2.2 Monodialectal Participants

Table B.5: Regression table for monodialectal participants in the rating task

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	0.64	0.11	0.38	35.00	5.72	<0.001	***
Cond_forward1	-0.98	0.15	-0.48	50.83	-6.42	<0.001	***
Cond_forward2	0.93	0.12	0.46	70.61	8.04	<0.001	***
Cond_forward3	0.63	0.10	0.31	111.47	6.63	<0.001	***
GuiseMAE	-1.54	0.17	-0.76	35.00	-9.20	<0.001	***
Cond_forward1:GuiseMAE	4.07	0.13	2.00	3272.38	32.18	<0.001	***
Cond_forward2:GuiseMAE	-0.63	0.13	-0.31	3272.48	-4.97	<0.001	***
Cond_forward3:GuiseMAE	-0.49	0.13	-0.24	3272.48	-3.90	<0.001	***

Table B.6: Post hoc pairwise comparisons for monodialectal participants by grammaticality condition and guise in the rating task

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
AAL Grammatical-Both	MAE Grammatical-Both	-1.07	0.18	51.47	-5.79	0.001
AAL Grammatical-Both	AAL Grammatical-AAL	-0.98	0.15	50.83	-6.42	0.001
AAL Grammatical-Both	MAE Grammatical-AAL	2.02	0.22	38.14	9.25	0.001
AAL Grammatical-Both	AAL Ungrammatical-Agreement	-0.05	0.20	43.72	-0.25	1

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
AAL Grammatical-Both	MAE Ungrammatical-Agreement	2.32	0.22	37.96	10.36	0.001
AAL Grammatical-Both	AAL Ungrammatical-Other	0.58	0.21	42.10	2.74	0.14
AAL Grammatical-Both	MAE Ungrammatical-Other	2.46	0.23	37.85	10.79	0.001
MAE Grammatical-Both	AAL Grammatical-AAL	0.09	0.23	37.87	0.38	1
MAE Grammatical-Both	MAE Grammatical-AAL	3.09	0.15	50.83	20.14	0.001
MAE Grammatical-Both	AAL Ungrammatical-Agreement	1.02	0.28	36.87	3.65	0.02
MAE Grammatical-Both	MAE Ungrammatical-Agreement	3.39	0.20	43.69	17.38	0.001
MAE Grammatical-Both	AAL Ungrammatical-Other	1.65	0.30	36.59	5.48	0.001
MAE Grammatical-Both	MAE Ungrammatical-Other	3.53	0.21	42.10	16.59	0.001
AAL Grammatical-AAL	MAE Grammatical-AAL	3.00	0.18	51.47	16.23	0.001
AAL Grammatical-AAL	AAL Ungrammatical-Agreement	0.93	0.12	70.61	8.04	0.001
AAL Grammatical-AAL	MAE Ungrammatical-Agreement	3.31	0.17	40.63	19.81	0.001
AAL Grammatical-AAL	AAL Ungrammatical-Other	1.57	0.14	56.92	11.52	0.001
AAL Grammatical-AAL	MAE Ungrammatical-Other	3.44	0.17	40.72	20.80	0.001
MAE Grammatical-AAL	AAL Ungrammatical-Agreement	-2.07	0.23	37.89	-9.10	0.001
MAE Grammatical-AAL	MAE Ungrammatical-Agreement	0.31	0.12	70.51	2.63	0.16
MAE Grammatical-AAL	AAL Ungrammatical-Other	-1.43	0.25	37.37	-5.76	0.001
MAE Grammatical-AAL	MAE Ungrammatical-Other	0.44	0.14	56.92	3.27	0.04
AAL Ungrammatical-Agreement	MAE Ungrammatical-Agreement	2.37	0.18	51.50	12.83	0.001
AAL Ungrammatical-Agreement	AAL Ungrammatical-Other	0.63	0.10	111.47	6.63	0.001
AAL Ungrammatical-Agreement	MAE Ungrammatical-Other	2.51	0.17	40.23	14.53	0.001
MAE Ungrammatical-Agreement	AAL Ungrammatical-Other	-1.74	0.20	38.71	-8.62	0.001
MAE Ungrammatical-Agreement	MAE Ungrammatical-Other	0.14	0.10	111.24	1.46	0.83

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
AAL Ungrammatical-Other	MAE Ungrammatical-Other	1.88	0.18	51.47	10.16	0.001

B.1.2.3 Comparing Groups

Table B.7: Regression table for comparison between groups

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	-0.32	0.09	0.00	68.28	-3.56	<0.001	***
Group_eff	0.05	0.17	0.01	61.01	0.28	0.78	
Cond_eff	0.95	0.05	0.21	2860.93	19.40	<0.001	***
Guise_eff	-2.46	0.18	-0.59	60.97	-13.51	<0.001	***
Group_eff:Cond_eff	-0.62	0.10	-0.07	2822.00	-6.50	<0.001	***
Group_eff:Guise_eff	-0.44	0.36	-0.05	60.88	-1.20	0.236	
Cond_eff:Guise_eff	-1.09	0.10	-0.12	2863.68	-11.40	<0.001	***
Group_eff:Cond_eff:Guise_eff	0.93	0.19	0.05	2847.54	4.85	<0.001	***

B.1.2.4 Phenomena

Null copula

Table B.8: Regression table for bidialectal participants, null copula items

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	1.35	0.19	0.65	72.38	6.98	<0.001	***
GuiseMAE	0.38	0.26	0.17	73.55	1.45	0.152	
CondGrammatical-AAL	0.50	0.19	0.23	989.81	2.69	0.007	**
CondMarginal-AAL	0.27	0.19	0.12	995.10	1.44	0.149	
CondUngrammatical-Agreement	-1.51	0.19	-0.69	986.23	-8.12	<0.001	***
CondUngrammatical-Other	-2.35	0.19	-1.08	986.73	-12.59	<0.001	***
GuiseMAE:CondGrammatical-AAL	-3.32	0.26	-1.52	992.74	-12.56	<0.001	***
GuiseMAE:CondMarginal-AAL	-3.07	0.26	-1.41	997.56	-11.64	<0.001	***
GuiseMAE:CondUngrammatical-Agreement	-1.97	0.26	-0.90	980.90	-7.50	<0.001	***
GuiseMAE:CondUngrammatical-Other	-1.47	0.26	-0.67	987.46	-5.58	<0.001	***

Table B.9: Post hoc pairwise comparisons for bidialectal participants for null copula items by grammaticality condition and guise.

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
AAL Grammatical-Both	MAE Grammatical-Both	-0.38	0.26	73.55	-1.45	0.91
AAL Grammatical-Both	AAL Grammatical-AAL	-0.50	0.19	989.81	-2.69	0.18
AAL Grammatical-Both	MAE Grammatical-AAL	2.44	0.26	73.80	9.37	0.001
AAL Grammatical-Both	AAL Marginal-AAL	-0.27	0.19	995.10	-1.44	0.91
AAL Grammatical-Both	MAE Marginal-AAL	2.43	0.26	73.45	9.33	0.001
AAL Grammatical-Both	AAL Ungrammatical-Agreement	1.51	0.19	986.23	8.12	0.001
AAL Grammatical-Both	MAE Ungrammatical-Agreement	3.11	0.26	73.70	11.95	0.001
AAL Grammatical-Both	AAL Ungrammatical-Other	2.35	0.19	986.73	12.59	0.001
AAL Grammatical-Both	MAE Ungrammatical-Other	3.44	0.26	73.52	13.22	0.001
MAE Grammatical-Both	AAL Grammatical-AAL	-0.13	0.26	73.43	-0.48	1
MAE Grammatical-Both	MAE Grammatical-AAL	2.82	0.19	988.77	15.10	0.001
MAE Grammatical-Both	AAL Marginal-AAL	0.11	0.26	73.08	0.41	1
MAE Grammatical-Both	MAE Marginal-AAL	2.80	0.19	994.17	15.04	0.001
MAE Grammatical-Both	AAL Ungrammatical-Agreement	1.89	0.26	73.33	7.27	0.001
MAE Grammatical-Both	MAE Ungrammatical-Agreement	3.49	0.19	986.11	18.75	0.001
MAE Grammatical-Both	AAL Ungrammatical-Other	2.72	0.26	73.16	10.49	0.001
MAE Grammatical-Both	MAE Ungrammatical-Other	3.82	0.19	986.31	20.52	0.001
AAL Grammatical-AAL	MAE Grammatical-AAL	2.94	0.26	73.90	11.30	0.001
AAL Grammatical-AAL	AAL Marginal-AAL	0.23	0.19	989.57	1.25	0.96
AAL Grammatical-AAL	MAE Marginal-AAL	2.93	0.26	73.45	11.26	0.001
AAL Grammatical-AAL	AAL Ungrammatical-Agreement	2.02	0.19	995.31	10.80	0.001
AAL Grammatical-AAL	MAE Ungrammatical-Agreement	3.61	0.26	73.44	13.89	0.001
AAL Grammatical-AAL	AAL Ungrammatical-Other	2.85	0.19	986.26	15.29	0.001
AAL Grammatical-AAL	MAE Ungrammatical-Other	3.94	0.26	73.70	15.15	0.001

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
MAE Grammatical-AAL	AAL Marginal-AAL	-2.71	0.26	73.42	-10.42	0.001
MAE Grammatical-AAL	MAE Marginal-AAL	-0.01	0.19	988.13	-0.08	1
MAE Grammatical-AAL	AAL Ungrammatical-Agreement	-0.93	0.26	73.41	-3.56	0.02
MAE Grammatical-AAL	MAE Ungrammatical-Agreement	0.67	0.19	993.76	3.59	0.01
MAE Grammatical-AAL	AAL Ungrammatical-Other	-0.09	0.26	73.67	-0.35	1
MAE Grammatical-AAL	MAE Ungrammatical-Other	1.00	0.19	986.39	5.36	0.001
AAL Marginal-AAL	MAE Marginal-AAL	2.70	0.26	73.17	10.37	0.001
AAL Marginal-AAL	AAL Ungrammatical-Agreement	1.78	0.19	989.06	9.59	0.001
AAL Marginal-AAL	MAE Ungrammatical-Agreement	3.38	0.26	73.08	13.02	0.001
AAL Marginal-AAL	AAL Ungrammatical-Other	2.62	0.19	994.28	14.05	0.001
AAL Marginal-AAL	MAE Ungrammatical-Other	3.71	0.26	73.08	14.28	0.001
MAE Marginal-AAL	AAL Ungrammatical-Agreement	-0.91	0.26	73.08	-3.51	0.03
MAE Marginal-AAL	MAE Ungrammatical-Agreement	0.69	0.19	989.01	3.68	0.009
MAE Marginal-AAL	AAL Ungrammatical-Other	-0.08	0.26	73.08	-0.30	1
MAE Marginal-AAL	MAE Ungrammatical-Other	1.01	0.19	994.30	5.45	0.001
AAL Ungrammatical-Agreement	MAE Ungrammatical-Agreement	1.60	0.26	73.17	6.15	0.001
AAL Ungrammatical-Agreement	AAL Ungrammatical-Other	0.83	0.19	988.97	4.48	0.001
AAL Ungrammatical-Agreement	MAE Ungrammatical-Other	1.93	0.26	73.08	7.41	0.001
MAE Ungrammatical-Agreement	AAL Ungrammatical-Other	-0.76	0.26	73.08	-2.94	0.11
MAE Ungrammatical-Agreement	MAE Ungrammatical-Other	0.33	0.19	988.94	1.77	0.75
AAL Ungrammatical-Other	MAE Ungrammatical-Other	1.09	0.26	73.17	4.20	0.003

Table B.10: Regression table for monodialectal participants, null copula items

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	0.40	0.14	0.24	279.77	2.82	0.005	**
GuiseMAE	1.33	0.17	0.66	1395.00	7.78	<0.001	***
CondGrammatical-AAL	1.15	0.17	0.57	1395.00	6.72	<0.001	***
CondMarginal-AAL	0.94	0.17	0.47	1395.00	5.54	<0.001	***
CondUngrammatical-Agreement	0.16	0.17	0.08	1395.00	0.94	0.349	
CondUngrammatical-Other	-0.70	0.17	-0.35	1395.00	-4.12	<0.001	***
GuiseMAE:CondGrammatical-AAL	-4.23	0.24	-2.10	1395.00	-17.55	<0.001	***
GuiseMAE:CondMarginal-AAL	-3.63	0.24	-1.81	1395.00	-15.07	<0.001	***
GuiseMAE:CondUngrammatical-Agreement	-3.78	0.24	-1.88	1395.00	-15.68	<0.001	***
GuiseMAE:CondUngrammatical-Other	-2.99	0.24	-1.48	1395.00	-12.39	<0.001	***

Table B.11: Post hoc pairwise comparisons for monodialectal participants for null copula items by grammaticality condition and guise.

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
AAL Grammatical-Both	MAE Grammatical-Both	-1.33	0.17	1395	-7.78	0.001
AAL Grammatical-Both	AAL Grammatical-AAL	-1.15	0.17	1395	-6.72	0.001
AAL Grammatical-Both	MAE Grammatical-AAL	1.76	0.17	1395	10.31	0.001
AAL Grammatical-Both	AAL Marginal-AAL	-0.94	0.17	1395	-5.54	0.001
AAL Grammatical-Both	MAE Marginal-AAL	1.36	0.17	1395	7.99	0.001
AAL Grammatical-Both	AAL Ungrammatical-Agreement	-0.16	0.17	1395	-0.94	1
AAL Grammatical-Both	MAE Ungrammatical-Agreement	2.29	0.17	1395	13.45	0.001
AAL Grammatical-Both	AAL Ungrammatical-Other	0.70	0.17	1395	4.12	0.002
AAL Grammatical-Both	MAE Ungrammatical-Other	2.36	0.17	1395	13.86	0.001
MAE Grammatical-Both	AAL Grammatical-AAL	0.18	0.17	1395	1.06	0.99
MAE Grammatical-Both	MAE Grammatical-AAL	3.08	0.17	1395	18.10	0.001
MAE Grammatical-Both	AAL Marginal-AAL	0.38	0.17	1395	2.24	0.43
MAE Grammatical-Both	MAE Marginal-AAL	2.69	0.17	1395	15.77	0.001
MAE Grammatical-Both	AAL Ungrammatical-Agreement	1.17	0.17	1395	6.85	0.001
MAE Grammatical-Both	MAE Ungrammatical-Agreement	3.62	0.17	1395	21.23	0.001

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
MAE Grammatical-Both	AAL Ungrammatical-Other	2.03	0.17	1395	11.90	0.001
MAE Grammatical-Both	MAE Ungrammatical-Other	3.69	0.17	1395	21.64	0.001
AAL Grammatical-AAL	MAE Grammatical-AAL	2.90	0.17	1395	17.04	0.001
AAL Grammatical-AAL	AAL Marginal-AAL	0.20	0.17	1395	1.18	0.98
AAL Grammatical-AAL	MAE Marginal-AAL	2.51	0.17	1395	14.71	0.001
AAL Grammatical-AAL	AAL Ungrammatical-Agreement	0.99	0.17	1395	5.79	0.001
AAL Grammatical-AAL	MAE Ungrammatical-Agreement	3.44	0.17	1395	20.17	0.001
AAL Grammatical-AAL	AAL Ungrammatical-Other	1.85	0.17	1395	10.84	0.001
AAL Grammatical-AAL	MAE Ungrammatical-Other	3.51	0.17	1395	20.58	0.001
MAE Grammatical-AAL	AAL Marginal-AAL	-2.70	0.17	1395	-15.85	0.001
MAE Grammatical-AAL	MAE Marginal-AAL	-0.40	0.17	1395	-2.32	0.37
MAE Grammatical-AAL	AAL Ungrammatical-Agreement	-1.92	0.17	1395	-11.25	0.001
MAE Grammatical-AAL	MAE Ungrammatical-Agreement	0.53	0.17	1395	3.14	0.05
MAE Grammatical-AAL	AAL Ungrammatical-Other	-1.06	0.17	1395	-6.20	0.001
MAE Grammatical-AAL	MAE Ungrammatical-Other	0.60	0.17	1395	3.55	0.01
AAL Marginal-AAL	MAE Marginal-AAL	2.31	0.17	1395	13.53	0.001
AAL Marginal-AAL	AAL Ungrammatical-Agreement	0.78	0.17	1395	4.61	0.001
AAL Marginal-AAL	MAE Ungrammatical-Agreement	3.24	0.17	1395	18.99	0.001
AAL Marginal-AAL	AAL Ungrammatical-Other	1.65	0.17	1395	9.66	0.001
AAL Marginal-AAL	MAE Ungrammatical-Other	3.31	0.17	1395	19.40	0.001
MAE Marginal-AAL	AAL Ungrammatical-Agreement	-1.52	0.17	1395	-8.93	0.001
MAE Marginal-AAL	MAE Ungrammatical-Agreement	0.93	0.17	1395	5.46	0.001
MAE Marginal-AAL	AAL Ungrammatical-Other	-0.66	0.17	1395	-3.87	0.004
MAE Marginal-AAL	MAE Ungrammatical-Other	1.00	0.17	1395	5.87	0.001
AAL Ungrammatical-Agreement	MAE Ungrammatical-Agreement	2.45	0.17	1395	14.39	0.001
AAL Ungrammatical-Agreement	AAL Ungrammatical-Other	0.86	0.17	1395	5.05	0.001

Cell 1	Cell 2	estimate	SE	df	t.ratio	p.value
AAL Ungrammatical-Agreement	MAE Ungrammatical-Other	2.52	0.17	1395	14.79	0.001
MAE Ungrammatical-Agreement	AAL Ungrammatical-Other	-1.59	0.17	1395	-9.33	0.001
MAE Ungrammatical-Agreement	MAE Ungrammatical-Other	0.07	0.17	1395	0.41	1
AAL Ungrammatical-Other	MAE Ungrammatical-Other	1.66	0.17	1395	9.74	0.001

Subject-verb agreement

Table B.12: Regression table for bidialectal participants, SVA items

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	1.43	0.20	0.70	88.28	7.02	<0.001	***
GuiseMAE	-0.07	0.26	-0.03	115.55	-0.29	0.775	
CondGrammatical-AAL	0.22	0.20	0.10	1624.87	1.10	0.272	
CondUngrammatical-Agreement	-1.12	0.20	-0.51	1641.47	-5.55	<0.001	***
CondUngrammatical-Other	-2.00	0.20	-0.91	1640.05	-9.92	<0.001	***
PhenomenonWAS	0.01	0.21	0.01	1105.97	0.06	0.95	
GuiseMAE:CondGrammatical-AAL	-2.70	0.29	-1.23	1632.41	-9.48	<0.001	***
GuiseMAE:CondUngrammatical-Agreement	-1.85	0.29	-0.84	1628.01	-6.48	<0.001	***
GuiseMAE:CondUngrammatical-Other	-0.92	0.28	-0.42	1621.43	-3.24	0.001	**
GuiseMAE:PhenomenonWAS	0.18	0.29	0.08	1626.78	0.63	0.528	
CondGrammatical-AAL:PhenomenonWAS	0.19	0.28	0.09	1623.98	0.66	0.507	
CondUngrammatical-Agreement:PhenomenonWAS	-1.18	0.29	-0.54	1640.43	-4.15	<0.001	***
CondUngrammatical-Other:PhenomenonWAS	0.18	0.29	0.08	1639.17	0.62	0.532	
GuiseMAE:CondGrammatical-AAL:PhenomenonWAS	-0.75	0.40	-0.34	1631.90	-1.85	0.065	.
GuiseMAE:CondUngrammatical-Agreement:PhenomenonWAS	0.87	0.40	0.39	1626.81	2.15	0.032	*
GuiseMAE:CondUngrammatical-Other:PhenomenonWAS	-0.68	0.40	-0.31	1620.55	-1.68	0.094	.

Table B.13: Pairwise comparisons by phenomenon for bidialectal participants

cell1	cell2	estimate	SE	df	t.ratio	p
VS AAL Grammatical-Both	WAS AAL Grammatical-Both	-0.01	0.21	1105.97	-0.06	$p = 1$
VS MAE Grammatical-Both	WAS MAE Grammatical-Both	-0.19	0.21	1052.88	-0.93	$p = 1$
VS AAL Grammatical-AAL	WAS AAL Grammatical-AAL	-0.20	0.21	1066.79	-0.98	$p = 1$
VS MAE Grammatical-AAL	WAS MAE Grammatical-AAL	0.36	0.21	1048.81	1.76	$p =$ 0.93
VS AAL Ungrammatical-Agreement	WAS AAL Ungrammatical-Agreement	1.17	0.21	1027.74	5.65	$p <$ 0.001
VS MAE Ungrammatical-Agreement	WAS MAE Ungrammatical-Agreement	0.12	0.21	1026.13	0.59	$p = 1$
VS AAL Ungrammatical-Other	WAS AAL Ungrammatical-Other	-0.19	0.21	1013.16	-0.92	$p = 1$
VS MAE Ungrammatical-Other	WAS MAE Ungrammatical-Other	0.30	0.21	978.29	1.47	$p =$ 0.99

Table B.14: Regression table for monodialectal participants, SVA items

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	0.35	0.16	0.23	140.50	2.19	0.03	*
GuiseMAE	1.13	0.23	0.55	125.06	4.97	<0.001	***
CondGrammatical-AAL	0.91	0.17	0.45	2222.60	5.51	<0.001	***
CondUngrammatical-Agreement	0.49	0.17	0.24	2210.13	2.95	0.003	**
CondUngrammatical-Other	-0.30	0.17	-0.15	2221.33	-1.79	0.073	.
PhenomenonWAS	0.49	0.17	0.24	1245.16	2.89	0.004	**
GuiseMAE:CondGrammatical-AAL	-3.68	0.23	-1.80	2222.10	-15.70	<0.001	***
GuiseMAE:CondUngrammatical-Agreement	-3.52	0.23	-1.71	2222.42	-14.97	<0.001	***
GuiseMAE:CondUngrammatical-Other	-3.33	0.24	-1.62	2228.86	-14.15	<0.001	***
GuiseMAE:PhenomenonWAS	-0.37	0.23	-0.18	2230.18	-1.58	0.115	
CondGrammatical-AAL:PhenomenonWAS	-0.03	0.23	-0.02	2221.92	-0.14	0.888	
CondUngrammatical-Agreement:PhenomenonWAS	-0.98	0.23	-0.48	2209.97	-4.19	<0.001	***
CondUngrammatical-Other:PhenomenonWAS	-0.45	0.23	-0.22	2220.10	-1.93	0.054	.

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
GuiseMAE:CondGrammatical-AAL:PhenomenonWAS	-0.60	0.33	-0.29	2220.80	-1.81	0.07	.
GuiseMAE:CondUngrammatical-Agreement:PhenomenonWAS	0.47	0.33	0.23	2219.31	1.42	0.156	
GuiseMAE:CondUngrammatical-Other:PhenomenonWAS	0.79	0.33	0.38	2230.46	2.37	0.018	*

Table B.15: Pairwise comparisons by phenomenon for monodialectal participants

cell1	cell2	estimate	SE	df	t.ratio	p
VS AAL Grammatical-Both	WAS AAL Grammatical-Both	-0.49	0.17	1245.16	-2.89	$p =$ 0.22
VS MAE Grammatical-Both	WAS MAE Grammatical-Both	-0.12	0.17	1045.77	-0.70	$p = 1$
VS AAL Grammatical-AAL	WAS AAL Grammatical-AAL	-0.46	0.17	1184.67	-2.69	$p =$ 0.34
VS MAE Grammatical-AAL	WAS MAE Grammatical-AAL	0.52	0.17	993.93	3.02	$p =$ 0.16
VS AAL Ungrammatical-Agreement	WAS AAL Ungrammatical-Agreement	0.49	0.17	1132.27	2.89	$p =$ 0.22
VS MAE Ungrammatical-Agreement	WAS MAE Ungrammatical-Agreement	0.39	0.17	1100.08	2.30	$p =$ 0.62
VS AAL Ungrammatical-Other	WAS AAL Ungrammatical-Other	-0.04	0.17	1023.74	-0.21	$p = 1$
VS MAE Ungrammatical-Other	WAS MAE Ungrammatical-Other	-0.46	0.17	1069.12	-2.67	$p =$ 0.35

B.1.2.5 Continuous analysis

Table B.16: Regression table for rating task, comparing across levels of AAL usage

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	0.78	0.10	-0.03	130.67	8.14	<0.001	***
Cond_eff	1.77	0.16	0.50	148.97	11.40	<0.001	***
aal_self_contr1	0.34	0.19	0.19	111.33	1.74	0.085	.
aal_self_contr2	-0.08	0.19	-0.04	111.32	-0.39	0.696	
aal_self_contr3	0.14	0.31	0.08	112.81	0.44	0.66	
Cond_eff:aal_self_contr1	-0.41	0.29	-0.11	110.44	-1.39	0.169	

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
Cond_eff:aal_self_contr2	-0.50	0.29	-0.14	110.33	-1.72	0.089	.
Cond_eff:aal_self_contr3	-0.46	0.48	-0.13	112.40	-0.95	0.342	

B.2 Transcription task

B.2.1 Materials

B.2.1.1 Critical Items

Verbal (S)

1. He commit some crimes
2. He create space for that
3. The girl create something pretty.
4. The child get so mad
5. He get scared all the time
6. My mom keep stuff up there.
7. The dog keep sitting right here
8. He speak Spanish
9. He take so long to get ready
10. He take so much for granted
11. She take some time
12. She speak several languages
13. He break some laws
14. She break some of the things away
15. She comfort some families
16. He sit sometimes in the park
17. She visit sometimes
18. She seek some new opportunities there

(WAS)

1. Last week, they was definitely fighting
2. The roads was too slick last night.
3. Things was pretty hot during the war
4. Things was cheaper back then.
5. At that point, my friends was from the area
6. Our lives was great in those days.
7. My parents was getting older then.
8. Years ago, the kids was pretty close.
9. Those relationships was just really close in those days.
10. We was just so taken aback.
11. The neighborhoods was close at one point.
12. They was funding the program for a while.
13. The parents was very much involved in the school for a while
14. You was twenty at the time
15. They was getting ready to build it that year

16. They was born in the city
17. Back in school, my other friends was in band.
18. I liked it when you was there

B.2.1.2 Fillers

Negative Inversion

1. • **AAL:** Can't nobody stop you from trying
   • **MAE:** Nobody can stop you from trying
2. • **AAL:** Can't nobody afford that
   • **MAE:** Nobody can afford that
3. • **AAL:** Don't nobody really wanna volunteer no more
   • **MAE:** Nobody wanted to volunteer anymore
4. • **AAL:** Can't nobody make nobody do nothing
   • **MAE:** Nobody can make anyone do anything
5. • **AAL:** Can't nobody take that away from you
   • **MAE:** Nobody can take that away from you
6. • **AAL:** Ain't nobody gonna bother you there
   • **MAE:** Nobody will bother you there
7. • **AAL:** Don't nobody wanna hear what you gotta say
   • **MAE:** Nobody would hear what you have to say
8. • **AAL:** Don't nobody even know that this is exercise
   • **MAE:** Nobody would know that this is exercise
9. • **AAL:** Didn't nobody believe I was hurt
   • **MAE:** Nobody believed I was hurt
10. • **AAL:** Didn't nobody hardly stay home then
    • **MAE:** Nobody stayed home then.
11. • **AAL:** Ain't nobody gonna bother you there
    • **MAE:** Nobody will bother you here
12. • **AAL:** Ain't nobody else doing it
    • **MAE:** Nobody else will do it
13. • **AAL:** Ain't nobody gonna survive that long
    • **MAE:** Nobody would survive that long
14. • **AAL:** Don't nobody want to be from there
    • **MAE:** Nobody would want to be from there
15. • **AAL:** Ain't nobody about to change that
    • **MAE:** Nobody will change that
16. • **AAL:** Don't nobody know that feeling

- **MAE:** Nobody would know that feeling

17. • **AAL:** Ain't nobody knew it
    • **MAE:** Nobody knew it

18. • **AAL:** Couldn't nobody go through that town
    • **MAE:** Nobody could go through that town

19. • **AAL:** Ain't nobody getting paid
    • **MAE:** Nobody was getting paid

20. • **AAL:** Ain't nobody here
    • **MAE:** Nobody is here

Negative Concord

1. They didn't bite {no/any} cookie
2. They shouldn't split {no/any} cookie
3. They don't want {no/any} garden
4. He didn't grow {no/any} garden
5. He won't write {no/any} letter
6. They won't sign {no/any} letter
7. That didn't cost {no/any} money
8. They didn't spend {no/any} money
9. They don't hear {no/any} music
10. They can't make {no/any} music
11. He couldn't taste {no/any} sugar
12. They shouldn't add {no/any} sugar
13. He didn't read {no/any} book
14. He didn't skim {no/any} book
15. She didn't have {no/any} pizza
16. They don't serve {no/any} pizza

B.2.2 Models

B.2.2.1 Bidialectal and Mondodialectal

Table B.17: Regression for transcription task with two groups

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-3.07	0.35	-3.19	-8.66	<0.001	***
Guise_eff	-2.69	0.60	-1.36	-4.47	<0.001	***
Group_eff	-0.99	0.49	-0.50	-2.00	0.046	*
ItemType_eff	1.04	0.32	0.54	3.27	0.001	**
Guise_eff:Group_eff	-0.68	0.68	-0.17	-1.00	0.317	
Guise_eff:ItemType_eff	-0.79	0.57	-0.19	-1.38	0.168	
Group_eff:ItemType_eff	0.82	0.46	0.20	1.77	0.078	.

Fixed Effect	Estimate	SE	Std. Estimate	z	p
Guise_eff:Group_eff:ItemType_eff	0.26	0.93	0.03	0.28	0.778

B.2.2.2 Continuous analysis

Table B.18: Regression for transcription task, comparing across levels of AAL usage

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-2.68	0.28	-2.70	-9.48	<0.001	***
Guise_eff	-2.84	0.46	-1.42	-6.22	<0.001	***
aal_self_fd1	-0.78	0.42	-0.78	-1.85	0.064	.
aal_self_fd2	0.22	0.40	0.22	0.54	0.586	
aal_self_fd3	-2.22	0.59	-2.22	-3.76	<0.001	***
Guise_eff:aal_self_fd1	-0.40	0.68	-0.20	-0.58	0.56	
Guise_eff:aal_self_fd2	-0.19	0.63	-0.09	-0.30	0.763	
Guise_eff:aal_self_fd3	-0.40	0.88	-0.20	-0.46	0.645	

Appendix C: Materials and models for Experiment 4

C.1 Stimuli

C.1.1 Visual world paradigm

C.1.1.1 Training

1. She used to feed the ducks
2. They like the infant
3. The pilots nap between flights
4. They couldn't find an acrobat
5. They hired a cook
6. I don't see a snake
7. They had a big cat
8. The rabbit can jump

C.1.1.2 Verb morphology

Note: all of these sentences are displayed with a plural marker on the noun, as this is how they appeared on the recording list, but the intent was for this to be ambiguous.

1. The acrobats sell out the venue
2. The acrobats sort the equipment
3. The acrobats satisfy an audience
4. The acrobats support a large ball
5. The cats sit in the shade
6. The cats stand far away
7. The cats start down the path
8. The cats stink all the time
9. The cooks sell a fancy meal
10. The cooks sort the fruit
11. The cooks stir a special sauce
12. The cooks satisfy a food blogger
13. The ducks sink in the water
14. The ducks start to leave
15. The ducks stay in the river
16. The ducks swim in the pond
17. The infants sound very happy
18. The infants sing with joy
19. The infants sleep in the bedroom
20. The infants smear all the food
21. The pilots salute the flag

22. The pilots sound a bit worried
23. The pilots say very little
24. The pilots support the decision
25. The rabbits sit every day
26. The rabbits snort on the ground
27. The rabbits snore at night
28. The rabbits suffer all the time
29. The snakes see in the dark
30. The snakes swim in the water
31. The snakes slither across the road
32. The snakes suffer every night

C.1.1.3 Noun morphology

1. The barber(s) can gossip all day
2. The barber(s) can cut a lot of hair
3. The barber(s) might like that new fade
4. The barber(s) couldn't copy that style
5. The bear(s) might hide in a cave
6. The bear(s) can bite the raw meat
7. The bear(s) did dent the trash can
8. The bear(s) couldn't grab fish from the river
9. The bird(s) did peck at the tree
10. The bird(s) couldn't eat from the feeder
11. The bird(s) can make a big nest
12. The bird(s) might want some food
13. The carpenter(s) did drive this huge truck
14. The carpenter(s) did carve a cool design
15. The carpenter(s) can't help with that
16. The carpenter(s) can build amazing furniture
17. The doctor(s) did charge so much money
18. The doctor(s) can't prevent that from happening
19. The doctor(s) did visit every patient
20. The doctor(s) might recommend surgery
21. The farmer(s) did harvest the corn
22. The farmer(s) might buy some equipment
23. The farmer(s) couldn't plant the wheat
24. The farmer(s) could paint some furniture
25. The lawyer(s) did travel so far
26. The lawyer(s) can't trick the jury
27. The lawyer(s) could plan a new strategy
28. The lawyer(s) might prefer a big payment
29. The lion(s) can't bite through everything
30. The lion(s) could hide in the tall grass
31. The lion(s) did attack some buffalo
32. The lion(s) could track some prey
33. The monkey(s) can jump so high
34. The monkey(s) might take some fruit

35. The monkey(s) couldn't grip the thick branch
36. The monkey(s) might jump on the tourist
37. The pastor(s) can't joke during the service
38. The pastor(s) did confront the congregation
39. The pastor(s) can talk about religion
40. The pastor(s) did connect with the community
41. The soldier(s) couldn't defeat the enemy
42. The soldier(s) can confront a challenge
43. The soldier(s) did pack for a long journey
44. The soldier(s) could fight for the cause
45. The squirrel(s) might tap the window
46. The squirrel(s) can't blink that fast
47. The squirrel(s) did collect food for the winter
48. The squirrel(s) might run up a tree
49. The teacher(s) couldn't prepare during lunch
50. The teacher(s) did write on the chalkboard
51. The teacher(s) did record the lecture
52. The teacher(s) can help the class
53. The dog(s) would bark at everything
54. The dog(s) might hunt on the trail
55. The dog(s) would look so happy
56. The dog(s) can't escape that fence
57. The turtle(s) can make some noise
58. The turtle(s) can walk on the log
59. The turtle(s) might block the road
60. The turtle(s) can't sleep in the marsh
61. The zebra(s) can't kick too hard
62. The zebra(s) could bolt across the field
63. The zebra(s) did look so playful
64. The zebra(s) could drink some water

C.1.1.4 Singular fillers for verb morphology items

1. The acrobat can flip in the air
2. The acrobat can balance on one foot
3. The cat would purr so loud
4. The cat did sleep on the rug
5. The cook would work so late
6. The cook would wait on the timer
7. The duck might float on the surface
8. The duck did quack so loud
9. The infant would cry in the back seat
10. The infant would giggle with joy
11. The pilot would fly every day
12. The pilot might rest in the cockpit
13. The rabbit would hop in the grass
14. The rabbit could chew the pellet
15. The snake can move so fast

16. The snake might attack out of nowhere

C.1.1.5 Negative concord fillers

1. He {ain't/didn't} read {no/any} book.
2. He {ain't/didn't} skim {no/any} book.
3. You couldn't buy {no/any} book.
4. She wouldn't borrow {no/any} book.
5. He won't spill {no/any} coffee.
6. He {ain't/didn't} pour {no/any} coffee.
7. They don't drink {no/any} coffee.
8. She {ain't/didn't} make {no/any} coffee.
9. They {ain't/didn't} bite {no/any} cookie.
10. They shouldn't split {no/any} cookie.
11. She can't bake {no/any} cookie.
12. He {ain't/didn't} offer {no/any} cookie.
13. He wouldn't see {no/any} doctor.
14. He wouldn't call {no/any} doctor.
15. They don't visit {no/any} doctor.
16. She {ain't/didn't} ask {no/any} doctor.
17. They don't want {no/any} garden.
18. He {ain't/didn't} grow {no/any} garden.
19. She won't water {no/any} garden.
20. They don't want {no/any} garden.
21. They {ain't/didn't} wear {no/any} jacket.
22. She {ain't/didn't} sew {no/any} jacket.
23. She {ain't/didn't} buy {no/any} jacket.
24. He {ain't/didn't} bring {no/any} jacket.
25. He won't write {no/any} letter.
26. They won't sign {no/any} letter.
27. I {ain't/didn't} get {no/any} letter.
28. He won't send {no/any} letter.
29. That {ain't/didn't} cost {no/any} money.
30. They {ain't/didn't} spend {no/any} money.
31. She won't borrow {no/any} money.
32. They don't save {no/any} money.
33. They don't hear {no/any} music.
34. They can't make {no/any} music.
35. I can't write {no/any} music.
36. We won't share {no/any} music.
37. She {ain't/didn't} have {no/any} pizza.
38. They don't serve {no/any} pizza.
39. We {ain't/didn't} order {no/any} pizza.
40. They won't deliver {no/any} pizza.
41. She can't ride {no/any} scooter.
42. They don't own {no/any} scooter.
43. You shouldn't rent {no/any} scooter.
44. I can't fix {no/any} scooter.

45. He couldn't taste {no/any} sugar.
46. They shouldn't add {no/any} sugar.
47. He {ain't/didn't} scoop {no/any} sugar.
48. You shouldn't use {no/any} sugar.

C.1.1.6 Null copula fillers

1. He('s) ringing the bell.
2. She('s) hearing the bell.
3. He('s) towing the boat.
4. She('s) partying on a boat.
5. She('s) grabbing the broom.
6. She('s) putting away the broom.
7. He('s) eating chicken.
8. He('s) cooking chicken.
9. She('s) repairing the clock.
10. She('s) watching the clock.
11. She('s) learning the guitar.
12. He('s) trying the guitar.
13. He('s) wearing a helmet.
14. She('s) holding a helmet.
15. He('s) following on a horse.
16. She('s) petting the horse.
17. He('s) chopping with the knife.
18. She('s) using the knife.
19. She('s) hunting a mouse.
20. He('s) catching the mouse.
21. She('s) decorating with some paint.
22. He('s) checking out some paint.
23. She('s) writing with her pencil.
24. He('s) finding the pencil.
25. She('s) looking at the phone.
26. He('s) hanging up the phone.
27. He('s) avoiding the sun.
28. She('s) lying in the sun.
29. He('s) using the toilet.
30. She('s) flushing the toilet.
31. She('s) fixing the truck.
32. He('s) following the truck.
33. She('s) a/an acrobat.
34. He('s) a barber.
35. He('s) on a boat.
36. He('s) a carpenter.
37. She('s) a cook.
38. She('s) a doctor.
39. He('s) a farmer.
40. He('s) a/an infant.
41. She('s) a lawyer.

42. He('s) a pastor.
43. She('s) on the phone.
44. She('s) a pilot.
45. He('s) a soldier.
46. He('s) out in the sun.
47. She('s) a teacher.
48. She('s) in the truck.

C.1.2 Audio clips for rating task

Table C.1: AAL guise audio files in the sentence rating task (Exp. 4)

Sentence	Key Word	Audio Source
Ain't nobody gon' bother you there	nobody	VWP Speaker
They good people	good	VWP Speaker
I never test well	test	CORAAL
It done got worse	done	CORAAL
Can't nobody afford that	can't	VWP Speaker
It wasn't really no transition	no	CORAAL
She always fussin	always	VWP Speaker
I don't like nothing sweet on my chicken	nothing	CORAAL

Table C.2: MAE guise audio files in the sentence rating task (Exp. 4)

Sentence	Key Word	Audio Source
Nobody wanted to volunteer anymore	anymore	VWP Speaker
She's a good mother	a	VWP Speaker
Well, mine was totally west	west	SBC
You mean, you actually thought that we had seen this before?	before	SBC
Nobody can stop you from trying	nobody	VWP Speaker
I don't see any back in here	any	SBC
He's meeting his four friends	meeting	VWP Speaker
Now you can't do anything to this guy	anything	SBC

C.2 Model output

C.2.1 Picture selections

Table C.3: Regression table for Experiment 3 critical verb morphology items, picture selection task

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	-2.67	0.19	-2.65	-14.42	<0.001	***
Guise_eff	-0.81	0.20	-0.40	-4.11	<0.001	***
Group_eff	-1.23	0.26	-0.61	-4.71	<0.001	***
Guise_eff:Group_eff	-0.38	0.31	-0.09	-1.23	0.217	

Table C.4: Regression table for Experiment 3 noun morphology items, picture selection task

Fixed Effect	Estimate	SE	Std. Estimate	z	p	
(Intercept)	5.35	0.39	5.32	13.72	<0.001	***
Guise_eff	0.66	0.59	0.32	1.11	0.268	
Group_eff	1.23	0.45	0.61	2.74	0.006	**
Guise_eff:Group_eff	0.75	0.88	0.19	0.86	0.39	

C.2.2 RTs

Table C.5: Regression table for Experiment 3 critical verb morphology items, picture selection RTs

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	6.66	0.05	-0.01	103.45	125.68	<0.001	***
Guise_eff	0.04	0.04	0.00	178.64	1.14	0.254	
Group_eff	0.30	0.09	0.27	78.37	3.30	0.001	**
response_eff	0.23	0.04	0.10	3989.45	6.62	<0.001	***
Guise_eff:Group_eff	-0.13	0.07	-0.04	440.00	-1.83	0.068	.
Guise_eff:response_eff	0.11	0.07	0.03	3479.13	1.66	0.098	.
Group_eff:response_eff	-0.20	0.07	-0.05	3966.98	-2.88	0.004	**
Guise_eff:Group_eff:response_eff	-0.05	0.14	-0.01	3699.55	-0.37	0.709	

Table C.6: Regression table for Experiment 3 noun morphology items, picture selection RTs

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	6.50	0.08	-0.03	382.88	82.72	<0.001	***
Guise_eff	0.09	0.12	0.00	3795.59	0.78	0.436	
Group_eff	-0.53	0.16	-0.28	369.07	-3.40	<0.001	***

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
correct_eff	0.45	0.12	0.06	3788.02	3.73	<0.001	***
Guise_eff:Group_eff	-0.03	0.24	-0.01	3779.41	-0.11	0.913	
Guise_eff:correct_eff	0.21	0.24	0.01	3795.50	0.85	0.397	
Group_eff:correct_eff	-0.22	0.24	-0.01	3774.74	-0.91	0.361	
Guise_eff:Group_eff:correct_eff	-0.01	0.48	0.00	3779.48	-0.03	0.98	

C.2.3 Rating task

C.2.3.1 Subject-verb agreement items

Table C.7: Regression table for Experiment 3 sentence rating task

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	1.77	0.14	0.86	156.96	13.12	<0.001	***
GuiseMAE	-0.16	0.20	-0.08	162.17	-0.84	0.405	
CondGrammatical-AAL	0.30	0.12	0.14	4035.66	2.49	0.013	*
CondUngrammatical-Agreement	-2.01	0.12	-0.94	4043.98	-16.45	<0.001	***
CondUngrammatical-Other	-2.50	0.12	-1.18	4049.64	-20.50	<0.001	***
GroupMonodialectal	-0.95	0.19	-0.45	142.33	-5.03	<0.001	***
GuiseMAE:CondGrammatical-AAL	-2.42	0.17	-1.14	4029.55	-14.03	<0.001	***
GuiseMAE:CondUngrammatical-Agreement	-1.06	0.17	-0.50	4038.20	-6.15	<0.001	***
GuiseMAE:CondUngrammatical-Other	-0.92	0.17	-0.43	4043.64	-5.34	<0.001	***
GuiseMAE:GroupMonodialectal	1.11	0.27	0.52	140.59	4.15	<0.001	***
CondGrammatical-AAL:GroupMonodialectal	0.36	0.18	0.17	4029.47	2.05	0.04	*
CondUngrammatical-Agreement:GroupMonodialectal	1.18	0.18	0.55	4039.20	6.66	<0.001	***
CondUngrammatical-Other:GroupMonodialectal	1.01	0.18	0.48	4046.02	5.75	<0.001	***
GuiseMAE:CondGrammatical-AAL:GroupMonodialectal	-1.30	0.25	-0.61	4024.38	-5.20	<0.001	***
GuiseMAE:CondUngrammatical-Agreement:GroupMonodialectal	-1.53	0.25	-0.72	4033.57	-6.14	<0.001	***
GuiseMAE:CondUngrammatical-Other:GroupMonodialectal	-1.39	0.25	-0.66	4039.79	-5.59	<0.001	***

Estimated marginal means

1. (Grammatical-Both AAL Bidialectal) - (Grammatical-AAL AAL Bidialectal): EMM=-0.3, SE=0.12, $t(4035.66) = -2.49$, $p = 0.48$
2. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Agreement AAL Bidialectal): EMM=2.01, SE=0.12, $t(4043.98) = 16.45$, $p < 0.001$
3. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Other AAL Bidialectal): EMM=2.5, SE=0.12, $t(4049.64) = 20.5$, $p < 0.001$
4. (Grammatical-Both AAL Bidialectal) - (Grammatical-Both MAE Bidialectal): EMM=0.16, SE=0.2, $t(162.17) = 0.84$, $p = 1$
5. (Grammatical-Both AAL Bidialectal) - (Grammatical-AAL MAE Bidialectal): EMM=2.28, SE=0.2, $t(162.11) = 11.67$, $p < 0.001$
6. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Agreement MAE Bidialectal): EMM=3.23, SE=0.2, $t(162.11) = 16.54$, $p < 0.001$
7. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Other MAE Bidialectal): EMM=3.58, SE=0.2, $t(162.08) = 18.36$, $p < 0.001$
8. (Grammatical-Both AAL Bidialectal) - (Grammatical-Both AAL Monodialectal): EMM=0.95, SE=0.19, $t(142.33) = 5.03$, $p < 0.001$
9. (Grammatical-Both AAL Bidialectal) - (Grammatical-AAL AAL Monodialectal): EMM=0.29, SE=0.19, $t(142.46) = 1.51$, $p = 0.98$
10. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal): EMM=1.78, SE=0.19, $t(142.75) = 9.42$, $p < 0.001$
11. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Other AAL Monodialectal): EMM=2.44, SE=0.19, $t(142.23) = 12.9$, $p < 0.001$
12. (Grammatical-Both AAL Bidialectal) - (Grammatical-Both MAE Monodialectal): EMM=0, SE=0.21, $t(263.89) = 0$, $p = 1$
13. (Grammatical-Both AAL Bidialectal) - (Grammatical-AAL MAE Monodialectal): EMM=3.05, SE=0.21, $t(263.98) = 14.55$, $p < 0.001$
14. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal): EMM=3.42, SE=0.21, $t(264.08) = 16.34$, $p < 0.001$
15. (Grammatical-Both AAL Bidialectal) - (Ungrammatical-Other MAE Monodialectal): EMM=3.8, SE=0.21, $t(264.52) = 18.14$, $p < 0.001$
16. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Agreement AAL Bidialectal): EMM=2.31, SE=0.12, $t(4035.22) = 18.96$, $p < 0.001$
17. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Other AAL Bidialectal): EMM=2.8, SE=0.12, $t(4044.46) = 23$, $p < 0.001$
18. (Grammatical-AAL AAL Bidialectal) - (Grammatical-Both MAE Bidialectal): EMM=0.47, SE=0.2, $t(162.09) = 2.39$, $p = 0.56$
19. (Grammatical-AAL AAL Bidialectal) - (Grammatical-AAL MAE Bidialectal): EMM=2.58, SE=0.2, $t(162.18) = 13.22$, $p < 0.001$
20. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Agreement MAE Bidialectal): EMM=3.53, SE=0.2, $t(162.12) = 18.1$, $p < 0.001$
21. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Other MAE Bidialectal): EMM=3.89, SE=0.2, $t(162.12) = 19.92$, $p < 0.001$
22. (Grammatical-AAL AAL Bidialectal) - (Grammatical-Both AAL Monodialectal): EMM=1.26, SE=0.19, $t(142.46) = 6.64$, $p < 0.001$
23. (Grammatical-AAL AAL Bidialectal) - (Grammatical-AAL AAL Monodialectal): EMM=0.59, SE=0.19, $t(142.6) = 3.12$, $p = 0.14$
24. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal): EMM=2.09, SE=0.19, $t(142.71) = 11.02$, $p < 0.001$

25. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Other AAL Monodialectal):
    EMM=2.74, SE=0.19, $t(142.24) = 14.51$, $p < 0.001$
26. (Grammatical-AAL AAL Bidialectal) - (Grammatical-Both MAE Monodialectal):
    EMM=0.3, SE=0.21, $t(263.99) = 1.45$, $p = 0.99$
27. (Grammatical-AAL AAL Bidialectal) - (Grammatical-AAL MAE Monodialectal):
    EMM=3.35, SE=0.21, $t(263.89) = 16.01$, $p < 0.001$
28. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal):
    EMM=3.73, SE=0.21, $t(263.98) = 17.79$, $p < 0.001$
29. (Grammatical-AAL AAL Bidialectal) - (Ungrammatical-Other MAE Monodialectal):
    EMM=4.11, SE=0.21, $t(264.49) = 19.6$, $p < 0.001$
30. (Ungrammatical-Agreement AAL Bidialectal) - (Ungrammatical-Other AAL Bidialectal):
    EMM=0.49, SE=0.12, $t(4036.56) = 4.06$, $p = 0.005$
31. (Ungrammatical-Agreement AAL Bidialectal) - (Grammatical-Both MAE Bidialectal):
    EMM=-1.84, SE=0.2, $t(161.86) = -9.44$, $p < 0.001$
32. (Ungrammatical-Agreement AAL Bidialectal) - (Grammatical-AAL MAE Bidialectal):
    EMM=0.27, SE=0.2, $t(161.88) = 1.39$, $p = 0.99$
33. (Ungrammatical-Agreement AAL Bidialectal) - (Ungrammatical-Agreement MAE Bidialectal):
    EMM=1.22, SE=0.2, $t(161.97) = 6.27$, $p < 0.001$
34. (Ungrammatical-Agreement AAL Bidialectal) - (Ungrammatical-Other MAE Bidialectal):
    EMM=1.58, SE=0.2, $t(161.89) = 8.09$, $p < 0.001$
35. (Ungrammatical-Agreement AAL Bidialectal) - (Grammatical-Both AAL Monodialectal):
    EMM=-1.05, SE=0.19, $t(142.18) = -5.58$, $p < 0.001$
36. (Ungrammatical-Agreement AAL Bidialectal) - (Grammatical-AAL AAL Monodialectal):
    EMM=-1.72, SE=0.19, $t(142.52) = -9.1$, $p < 0.001$
37. (Ungrammatical-Agreement AAL Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal):
    EMM=-0.22, SE=0.19, $t(142.64) = -1.18$, $p = 1$
38. (Ungrammatical-Agreement AAL Bidialectal) - (Ungrammatical-Other AAL Monodialectal):
    EMM=0.43, SE=0.19, $t(141.99) = 2.29$, $p = 0.63$
39. (Ungrammatical-Agreement AAL Bidialectal) - (Grammatical-Both MAE Monodialectal):
    EMM=-2.01, SE=0.21, $t(263.68) = -9.58$, $p < 0.001$
40. (Ungrammatical-Agreement AAL Bidialectal) - (Grammatical-AAL MAE Monodialectal):
    EMM=1.04, SE=0.21, $t(263.67) = 4.98$, $p < 0.001$
41. (Ungrammatical-Agreement AAL Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal):
    EMM=1.42, SE=0.21, $t(263.57) = 6.76$, $p < 0.001$
42. (Ungrammatical-Agreement AAL Bidialectal) - (Ungrammatical-Other MAE Monodialectal):
    EMM=1.8, SE=0.21, $t(264.07) = 8.58$, $p < 0.001$
43. (Ungrammatical-Other AAL Bidialectal) - (Grammatical-Both MAE Bidialectal):
    EMM=-2.34, SE=0.2, $t(161.83) = -11.98$, $p < 0.001$
44. (Ungrammatical-Other AAL Bidialectal) - (Grammatical-AAL MAE Bidialectal):
    EMM=-0.22, SE=0.2, $t(161.87) = -1.14$, $p = 1$
45. (Ungrammatical-Other AAL Bidialectal) - (Ungrammatical-Agreement MAE Bidialectal):
    EMM=0.73, SE=0.2, $t(161.89) = 3.73$, $p = 0.02$
46. (Ungrammatical-Other AAL Bidialectal) - (Ungrammatical-Other MAE Bidialectal):

EMM=1.08, SE=0.2, $t(161.98) = 5.56$, $p < 0.001$

47. (Ungrammatical-Other AAL Bidialectal) - (Grammatical-Both AAL Monodialectal):
EMM=-1.55, SE=0.19, $t(141.99) = -8.2$, $p < 0.001$
48. (Ungrammatical-Other AAL Bidialectal) - (Grammatical-AAL AAL Monodialectal):
EMM=-2.21, SE=0.19, $t(142.44) = -11.71$, $p < 0.001$
49. (Ungrammatical-Other AAL Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal):
EMM=-0.72, SE=0.19, $t(142.77) = -3.8$, $p = 0.02$
50. (Ungrammatical-Other AAL Bidialectal) - (Ungrammatical-Other AAL Monodialectal):
EMM=-0.06, SE=0.19, $t(142.12) = -0.33$, $p = 1$
51. (Ungrammatical-Other AAL Bidialectal) - (Grammatical-Both MAE Monodialectal):
EMM=-2.5, SE=0.21, $t(263.68) = -11.94$, $p < 0.001$
52. (Ungrammatical-Other AAL Bidialectal) - (Grammatical-AAL MAE Monodialectal):
EMM=0.55, SE=0.21, $t(263.68) = 2.62$, $p = 0.39$
53. (Ungrammatical-Other AAL Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal):
EMM=0.92, SE=0.21, $t(263.66) = 4.4$, $p = 0.002$
54. (Ungrammatical-Other AAL Bidialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=1.3, SE=0.21, $t(263.98) = 6.22$, $p < 0.001$
55. (Grammatical-Both MAE Bidialectal) - (Grammatical-AAL MAE Bidialectal):
EMM=2.11, SE=0.12, $t(4025.07) = 17.35$, $p < 0.001$
56. (Grammatical-Both MAE Bidialectal) - (Ungrammatical-Agreement MAE Bidialectal):
EMM=3.07, SE=0.12, $t(4032.43) = 25.14$, $p < 0.001$
57. (Grammatical-Both MAE Bidialectal) - (Ungrammatical-Other MAE Bidialectal):
EMM=3.42, SE=0.12, $t(4038.66) = 28.04$, $p < 0.001$
58. (Grammatical-Both MAE Bidialectal) - (Grammatical-Both AAL Monodialectal):
EMM=0.79, SE=0.21, $t(272.05) = 3.78$, $p = 0.02$
59. (Grammatical-Both MAE Bidialectal) - (Grammatical-AAL AAL Monodialectal):
EMM=0.12, SE=0.21, $t(272.55) = 0.59$, $p = 1$
60. (Grammatical-Both MAE Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal):
EMM=1.62, SE=0.21, $t(273.04) = 7.76$, $p < 0.001$
61. (Grammatical-Both MAE Bidialectal) - (Ungrammatical-Other AAL Monodialectal):
EMM=2.27, SE=0.21, $t(272.28) = 10.91$, $p < 0.001$
62. (Grammatical-Both MAE Bidialectal) - (Grammatical-Both MAE Monodialectal):
EMM=-0.16, SE=0.21, $t(118.15) = -0.77$, $p = 1$
63. (Grammatical-Both MAE Bidialectal) - (Grammatical-AAL MAE Monodialectal):
EMM=2.89, SE=0.21, $t(118.01) = 13.68$, $p < 0.001$
64. (Grammatical-Both MAE Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal):
EMM=3.26, SE=0.21, $t(118.03) = 15.46$, $p < 0.001$
65. (Grammatical-Both MAE Bidialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=3.64, SE=0.21, $t(118.2) = 17.25$, $p < 0.001$
66. (Grammatical-AAL MAE Bidialectal) - (Ungrammatical-Agreement MAE Bidialectal):
EMM=0.95, SE=0.12, $t(4024.18) = 7.81$, $p < 0.001$
67. (Grammatical-AAL MAE Bidialectal) - (Ungrammatical-Other MAE Bidialectal):
EMM=1.31, SE=0.12, $t(4032.8) = 10.72$, $p < 0.001$
68. (Grammatical-AAL MAE Bidialectal) - (Grammatical-Both AAL Monodialectal):
EMM=-1.33, SE=0.21, $t(272.16) = -6.36$, $p < 0.001$
69. (Grammatical-AAL MAE Bidialectal) - (Grammatical-AAL AAL Monodialectal):

EMM=-1.99, SE=0.21, $t(272.45) = -9.55$, $p < 0.001$

70. (Grammatical-AAL MAE Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal):
EMM=-0.5, SE=0.21, $t(272.95) = -2.38$, $p = 0.57$
71. (Grammatical-AAL MAE Bidialectal) - (Ungrammatical-Other AAL Monodialectal):
EMM=0.16, SE=0.21, $t(272.24) = 0.77$, $p = 1$
72. (Grammatical-AAL MAE Bidialectal) - (Grammatical-Both MAE Monodialectal):
EMM=-2.28, SE=0.21, $t(118.27) = -10.8$, $p < 0.001$
73. (Grammatical-AAL MAE Bidialectal) - (Grammatical-AAL MAE Monodialectal):
EMM=0.77, SE=0.21, $t(118.16) = 3.65$, $p = 0.03$
74. (Grammatical-AAL MAE Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal):
EMM=1.14, SE=0.21, $t(118.01) = 5.43$, $p < 0.001$
75. (Grammatical-AAL MAE Bidialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=1.52, SE=0.21, $t(118.2) = 7.23$, $p < 0.001$
76. (Ungrammatical-Agreement MAE Bidialectal) - (Ungrammatical-Other MAE Bidialectal):
EMM=0.36, SE=0.12, $t(4025.67) = 2.92$, $p = 0.21$
77. (Ungrammatical-Agreement MAE Bidialectal) - (Grammatical-Both AAL Monodialectal):
EMM=-2.28, SE=0.21, $t(272.17) = -10.92$, $p < 0.001$
78. (Ungrammatical-Agreement MAE Bidialectal) - (Grammatical-AAL AAL Monodialectal):
EMM=-2.94, SE=0.21, $t(272.56) = -14.11$, $p < 0.001$
79. (Ungrammatical-Agreement MAE Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal):
EMM=-1.45, SE=0.21, $t(272.86) = -6.94$, $p < 0.001$
80. (Ungrammatical-Agreement MAE Bidialectal) - (Ungrammatical-Other AAL Monodialectal):
EMM=-0.79, SE=0.21, $t(272.16) = -3.79$, $p = 0.02$
81. (Ungrammatical-Agreement MAE Bidialectal) - (Grammatical-Both MAE Monodialectal):
EMM=-3.23, SE=0.21, $t(118.2) = -15.31$, $p < 0.001$
82. (Ungrammatical-Agreement MAE Bidialectal) - (Grammatical-AAL MAE Monodialectal):
EMM=-0.18, SE=0.21, $t(118.27) = -0.86$, $p = 1$
83. (Ungrammatical-Agreement MAE Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal):
EMM=0.19, SE=0.21, $t(118.15) = 0.92$, $p = 1$
84. (Ungrammatical-Agreement MAE Bidialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=0.57, SE=0.21, $t(118.19) = 2.72$, $p = 0.33$
85. (Ungrammatical-Other MAE Bidialectal) - (Grammatical-Both AAL Monodialectal):
EMM=-2.63, SE=0.21, $t(272.18) = -12.63$, $p < 0.001$
86. (Ungrammatical-Other MAE Bidialectal) - (Grammatical-AAL AAL Monodialectal):
EMM=-3.3, SE=0.21, $t(272.59) = -15.82$, $p < 0.001$
87. (Ungrammatical-Other MAE Bidialectal) - (Ungrammatical-Agreement AAL Monodialectal):
EMM=-1.8, SE=0.21, $t(272.97) = -8.64$, $p < 0.001$
88. (Ungrammatical-Other MAE Bidialectal) - (Ungrammatical-Other AAL Monodialectal):
EMM=-1.15, SE=0.21, $t(272.06) = -5.5$, $p < 0.001$
89. (Ungrammatical-Other MAE Bidialectal) - (Grammatical-Both MAE Monodialectal):
EMM=-3.58, SE=0.21, $t(118.01) = -17$, $p < 0.001$
90. (Ungrammatical-Other MAE Bidialectal) - (Grammatical-AAL MAE Monodialectal):
EMM=-0.54, SE=0.21, $t(118.21) = -2.54$, $p = 0.45$

91. (Ungrammatical-Other MAE Bidialectal) - (Ungrammatical-Agreement MAE Monodialectal):
    EMM=-0.16, SE=0.21, $t(118.28) = -0.77$, $p = 1$
92. (Ungrammatical-Other MAE Bidialectal) - (Ungrammatical-Other MAE Monodialectal):
    EMM=0.22, SE=0.21, $t(118.32) = 1.03$, $p = 1$
93. (Grammatical-Both AAL Monodialectal) - (Grammatical-AAL AAL Monodialectal):
    EMM=-0.67, SE=0.13, $t(4020.63) = -5.23$, $p < 0.001$
94. (Grammatical-Both AAL Monodialectal) - (Ungrammatical-Agreement AAL Monodialectal):
    EMM=0.83, SE=0.13, $t(4028.35) = 6.51$, $p < 0.001$
95. (Grammatical-Both AAL Monodialectal) - (Ungrammatical-Other AAL Monodialectal):
    EMM=1.49, SE=0.13, $t(4026.25) = 11.68$, $p < 0.001$
96. (Grammatical-Both AAL Monodialectal) - (Grammatical-Both MAE Monodialectal):
    EMM=-0.95, SE=0.2, $t(160.28) = -4.68$, $p < 0.001$
97. (Grammatical-Both AAL Monodialectal) - (Grammatical-AAL MAE Monodialectal):
    EMM=2.1, SE=0.2, $t(160.32) = 10.32$, $p < 0.001$
98. (Grammatical-Both AAL Monodialectal) - (Ungrammatical-Agreement MAE Monodialectal):
    EMM=2.47, SE=0.2, $t(160.33) = 12.16$, $p < 0.001$
99. (Grammatical-Both AAL Monodialectal) - (Ungrammatical-Other MAE Monodialectal):
    EMM=2.85, SE=0.2, $t(160.56) = 14.03$, $p < 0.001$
100. (Grammatical-AAL AAL Monodialectal) - (Ungrammatical-Agreement AAL Monodialectal):
    EMM=1.5, SE=0.13, $t(4020.41) = 11.73$, $p < 0.001$
101. (Grammatical-AAL AAL Monodialectal) - (Ungrammatical-Other AAL Monodialectal):
    EMM=2.15, SE=0.13, $t(4028.6) = 16.9$, $p < 0.001$
102. (Grammatical-AAL AAL Monodialectal) - (Grammatical-Both MAE Monodialectal):
    EMM=-0.29, SE=0.2, $t(160.57) = -1.4$, $p = 0.99$
103. (Grammatical-AAL AAL Monodialectal) - (Grammatical-AAL MAE Monodialectal):
    EMM=2.76, SE=0.2, $t(160.54) = 13.6$, $p < 0.001$
104. (Grammatical-AAL AAL Monodialectal) - (Ungrammatical-Agreement MAE Monodialectal):
    EMM=3.14, SE=0.2, $t(160.57) = 15.44$, $p < 0.001$
105. (Grammatical-AAL AAL Monodialectal) - (Ungrammatical-Other MAE Monodialectal):
    EMM=3.52, SE=0.2, $t(160.84) = 17.3$, $p < 0.001$
106. (Ungrammatical-Agreement AAL Monodialectal) - (Ungrammatical-Other AAL Monodialectal):
    EMM=0.66, SE=0.13, $t(4021.06) = 5.15$, $p < 0.001$
107. (Ungrammatical-Agreement AAL Monodialectal) - (Grammatical-Both MAE Monodialectal):
    EMM=-1.78, SE=0.2, $t(160.83) = -8.76$, $p < 0.001$
108. (Ungrammatical-Agreement AAL Monodialectal) - (Grammatical-AAL MAE Monodialectal):
    EMM=1.27, SE=0.2, $t(160.81) = 6.23$, $p < 0.001$
109. (Ungrammatical-Agreement AAL Monodialectal) - (Ungrammatical-Agreement MAE Monodialectal):
    EMM=1.64, SE=0.2, $t(160.78) = 8.07$, $p < 0.001$
110. (Ungrammatical-Agreement AAL Monodialectal) - (Ungrammatical-Other MAE Monodi-

alectal):
EMM=2.02, SE=0.2, $t(161.08) = 9.94$, $p < 0.001$

111. (Ungrammatical-Other AAL Monodialectal) - (Grammatical-Both MAE Monodialectal):
EMM=-2.44, SE=0.2, $t(160.3) = -12$, $p < 0.001$
112. (Ungrammatical-Other AAL Monodialectal) - (Grammatical-AAL MAE Monodialectal):
EMM=0.61, SE=0.2, $t(160.33) = 3.01$, $p = 0.18$
113. (Ungrammatical-Other AAL Monodialectal) - (Ungrammatical-Agreement MAE Monodialectal):
EMM=0.98, SE=0.2, $t(160.31) = 4.85$, $p < 0.001$
114. (Ungrammatical-Other AAL Monodialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=1.36, SE=0.2, $t(160.54) = 6.71$, $p < 0.001$
115. (Grammatical-Both MAE Monodialectal) - (Grammatical-AAL MAE Monodialectal):
EMM=3.05, SE=0.13, $t(4014.81) = 23.96$, $p < 0.001$
116. (Grammatical-Both MAE Monodialectal) - (Ungrammatical-Agreement MAE Monodialectal):
EMM=3.42, SE=0.13, $t(4019.7) = 26.88$, $p < 0.001$
117. (Grammatical-Both MAE Monodialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=3.8, SE=0.13, $t(4019.76) = 29.84$, $p < 0.001$
118. (Grammatical-AAL MAE Monodialectal) - (Ungrammatical-Agreement MAE Monodialectal):
EMM=0.37, SE=0.13, $t(4014.86) = 2.94$, $p = 0.2$
119. (Grammatical-AAL MAE Monodialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=0.75, SE=0.13, $t(4020.53) = 5.91$, $p < 0.001$
120. (Ungrammatical-Agreement MAE Monodialectal) - (Ungrammatical-Other MAE Monodialectal):
EMM=0.38, SE=0.13, $t(4015.14) = 2.98$, $p = 0.18$

Table C.8: Regression table for Experiment 3 sentence rating task, comparing only Grammatical-AAL and Ungrammatical-Agreement conditions

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	-0.20	0.09	0.00	82.06	-2.10	0.039	*
Cond_eff	-1.27	0.11	-0.31	75.17	-11.89	<0.001	***
Guise_eff	-2.05	0.18	-0.49	73.39	-11.50	<0.001	***
Group_eff	-0.33	0.17	-0.08	63.61	-1.95	0.055	.
Cond_eff:Guise_eff	1.28	0.11	0.15	1844.28	11.47	<0.001	***
Cond_eff:Group_eff	0.67	0.19	0.08	63.11	3.53	<0.001	***
Guise_eff:Group_eff	-0.33	0.34	-0.04	64.01	-0.98	0.33	
Cond_eff:Guise_eff:Group_eff	-0.31	0.22	-0.02	1845.87	-1.39	0.165	

C.2.3.2 Habitual *be* items

Table C.9: Regression table for Experiment 3 sentence rating task, habitual *be* items

Fixed Effect	Estimate	SE	Std. Estimate	df	t	p	
(Intercept)	1.28	0.24	0.08	39.29	5.33	<0.001	***
GroupBidialectal	-1.23	0.32	-0.75	55.14	-3.89	<0.001	***
Cond2Correct	0.17	0.22	0.10	28.50	0.74	0.464	
GroupBidialectal:Cond2Correct	1.64	0.27	1.00	65.00	6.11	<0.001	***

Bibliography

Adler, N. E., Epel, E. S., Castellazzo, G., and Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, white women. *Health Psychology*, 19(6):586–592.

Austen, M. (2020). *The Role of Listener Experience in Perception of Conditioned Dialect Variation*. PhD thesis, The Ohio State University.

Baker-Bell, A. (2020). *Linguistic justice: Black language, literacy, identity, and pedagogy*. Routledge.

Ball, A. F. (1995). Language, learning, and linguistic competence of African American children: Torrey revisited. *Linguistics and Education*, 7(1):23–46.

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

Baugh, J. (1979). *Linguistic style-shifting in Black English*. PhD thesis, University of Pennsylvania.

Baugh, J. (1990). A Survey of the Suffix /-s/ Analyses in Black English. In Edmondson, J. A., Feagin, C., and Mühlhäusler, P., editors, *Development and Diversity: Language Variation across Time and Space: A Festschrift for Charles-James N. Bailey*, pages 297–307. The Summer Institute of Linguistics.

Beatty-Martínez, A. L., Guzzardo Tamargo, R. E., and Dussias, P. E. (2021). Phasic pupillary responses reveal differential engagement of attentional control in bilingual spoken language processing. *Scientific Reports*, 11(1):1–12.

Bender, E. M. (2000). *Syntactic variation and linguistic competence: The case of AAVE copula absence*. PhD thesis, Stanford University.

Bender, E. M. (2005). On the boundaries of linguistic competence: Matched-guise experiments as evidence of knowledge of grammar. *Lingua*, 115(11 SPEC. ISS.):1579–1598.

Beyer, T., Edwards, K. A., and Fuller, C. C. (2015). Misinterpretation of African American English BIN by adult speakers of Standard American English. *Language and Communication*, 45:59–69.

Bice, K. and Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language*, 196(July):104644.

Blanchette, F., Flannery, E., Jackson, C., and Reed, P. (2023). Adaptation at the syntax–semantics interface: Evidence from a vernacular structure. *Language and Speech*.

Blanchette, F. and Lukyanenko, C. (2019a). Asymmetries in the Acceptability and Felicity of English Negative Dependencies: Where Negative Concord and Negative Polarity (Do Not) Overlap. *Frontiers in Psychology*, 10(November):1–15.

Blanchette, F. and Lukyanenko, C. (2019b). Unacceptable grammars? An eye-tracking study of English negative concord. *Language and Cognition*, 11:1–40.

Blodgett, S. L., Green, L., and O'Connor, B. (2016). Demographic dialectal variation in social media: A case study of African-American English. In *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing*, pages 1119–1130.

Bock, K. and Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23(1):45–93.

Buson, L., Nardy, A., Muller, D., and Chevrot, J. P. (2018). The Sociolinguistic Repetition Task: A New Paradigm for Exploring the Cognitive Coherence of Language Varieties. *Topics in Cognitive Science*, 10(4):803–817.

Byrd, A. S., Huang, Y. T., and Edwards, J. (2023). The impact of dialect differences on spoken language comprehension. *Applied Psycholinguistics*, pages 1–24.

Caffarra, S. and Martin, C. D. (2019). Not all errors are the same: ERP sensitivity to error typicality in foreign accented speech perception. *Cortex*, 116:308–320.

Campbell-Kibler, K. (2011). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change*, 22(3):423–441.

Campbell-Kibler, K. (2016). Towards a cognitively realistic model of meaningful sociolinguistic variation. In Babel, A. M., editor, *Awareness and Control in Sociolinguistic Research*, pages 123–151. Cambridge University Press.

Charity, A. H., Scarborough, H. S., and Griffin, D. M. (2004). Familiarity with School English in African American Children and Its Relation to Early Reading Achievement. *Child Development*, 75(5):1340–1356.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, 69(5):817–828.

Cleveland, L. H. and Oetting, J. B. (2013). Children's Marking of Verbal –s by Nonmainstream English Dialect and Clinical Status. *American Journal of Speech-Language Pathology*, 22(4):604–614.

Cross, J. B., DeVaney, T., and Jones, G. (2001). Pre-Service Teacher Attitudes Toward Differing Dialects. *Linguistics and Education*, 12(2):211–227.

Dahan, D., Drucker, S. J., and Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108(3):710–718.

de Villiers, J. G. and Johnson, V. E. (2007). The information in third-person /s/: acquisition across dialects of American English. *Journal of Child Language*, 34(01):133.

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., and Martey, N. (2005). Santa Barabara corpus of spoken American English, Parts 1-4.

Eberhardt, M. and Freeman, K. (2015). 'First things first, I'm the realest': Linguistic appropriation, white privilege, and the hip-hop persona of Iggy Azalea. *Journal of Sociolinguistics*, 19(3):303–327.

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

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, (41):87–100.

Eckert, P. (2019). The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language*, 95(4):751–776.

Edwards, J., Gross, M., Chen, J., MacDonald, M. C., Kaplan, D., Brown, M., and Seidenberg, M. S. (2014). Dialect Awareness and Lexical Comprehension of Mainstream American English in African American English-Speaking Children. *Journal of Speech, Language, and Hearing Research*, 57(5):1883–1895.

Engel, A. and Hanulíková, A. (2020). Speaking Style Modulates Morphosyntactic Expectations in Young and Older Adults: Evidence from a Sentence Repetition Task. *Discourse Processes*, 57(9):749–769.

Ferreira, F., Bailey, K. G., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1):11–15.

Filson, N. A. (2019). *Exploring English Language Arts Pre-Service Teachers' Standard Language Ideologies: A Mixed Methods Study.* PhD thesis, North Carolina State University.

Garcia, F. M., Shen, G., Avery, T., Green, H. L., Godoy, P., Khamis, R., and Froud, K. (2022). Bidialectal and monodialectal differences in morphosyntactic processing of AAE and MAE: Evidence from ERPs and acceptability judgments. *Journal of Communication Disorders*, 100:106267.

Gatlin, B. and Wanzek, J. (2015). Relations Among Children's Use of Dialect and Literacy Skills: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, 58:1306–1318.

Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(20):8051–8056.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., and Fedorenko, E. (2017). Don't Underestimate the Benefits of Being Misunderstood. *Psychological Science*, 28(6):703–712.

Giles, H. and Ogay, T. (2007). Communication Accomodation Theory. In Whaley, B. B. and Samter, W., editors, *Explaining communication: Contemporary theories and exemplars*, pages 293–310. Mahwah, NJ.

Godley, A. J., Reaser, J., and Moore, K. G. (2015). Pre-service English Language Arts teachers' development of Critical Language Awareness for teaching. *Linguistics and Education*, 32:41–54.

Green, L. J. (2002). *African American English: a linguistic introduction.* Cambridge University Press, New York.

Green, L. J. (2011). *Language and the African American Child*. Cambridge University Press.

Green, L. J. (2019). Not All Zeros Are Not Equal in African American English. In Lightfoot, D. W. and Havenhill, J., editors, *Variable Properties in Language: Their Nature and Acquisition*, pages 183–193. Georgetown University Press.

Grey, S. and Van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, 42:93–108.

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

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

Hahne, A. and Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11(2):194–205.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., and Weber, A. (2012). When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *Journal of Cognitive Neuroscience*, 24:878–887.

Holliday, N. and Tano, M. (2021). "It's a Whole Vibe": testing evaluations of grammatical and ungrammatical AAE on Twitter. *Linguistics Vanguard*, 7(1).

Huang, Y. and Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition*, 200(December 2017):104251.

Jackson, J. E. (1998). *Linguistic aspect in African-American English-speaking children: An investigation of aspectual "be"*. PhD thesis, University of Massachusetts Amherst.

Johnson, V. E., De Villiers, J. G., and Seymour, H. N. (2005). Agreement without understanding? the case of third person singular /s/. *First Language*, 25(3):317–330.

Jones, T. (2015). Toward a description of African American Vernacular English dialect regions using "Black Twitter". *American Speech*, 90(4):403–440.

Jones, T., Kalbfeld, J. R., Hancock, R., Clark, R., and Jones, T. (2019). Testifying while black: An experimental study of court reporter accuracy in transcription of African American English. *Language*, 95(2).

Kaan, E., Kheder, S., Kreidler, A., Tomić, A., and Valdés Kroff, J. R. (2020). Processing Code-Switches in the Presence of Others: An ERP Study. *Frontiers in Psychology*, 11(June):1–18.

Kaschak, M. P. (2006). What this construction needs is generalized. *Memory and Cognition*, 34(2):368–379.

Kaschak, M. P. and Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133(3):450–467.

Kendall, T. and Farrington, C. (2021). The Corpus of Regional African American Language.

Kim, A. and Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, 118(1-2):15–22.

King, S. (2020). From African American Vernacular English to African American Language: Rethinking the Study of Race and Language in African Americans' Speech. *Annual Review of Linguistics*, 6:285–300.

King, S. and Sumner, M. (2014). Voices and Variants: Effects of Voice on the Form-Based Processing of Words with Different Phonological Variants. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Kirk, N. W., Kempe, V., Scott-Brown, K. C., Philipp, A., and Declerck, M. (2018). Can monolinguals be like bilinguals? Evidence from dialect switching. *Cognition*, 170:164–178.

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

Labov, W. (1969). Contraction, Deletion, and Inherent Variability of the English Copula. *Language*, 45(4):715–762.

Labov, W. (1973). Where do grammars stop? In Shuy, R., editor, *Report of the Twenty-Third Annual Round Table Meeting on Linguistics and Language Studies*, pages 43–88. Georgetown University Press, Washington, DC.

Labov, W. (2012). *Dialect diversity in America: The politics of language change*. University of Virginia Press.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., and Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4):431–463.

Labov, W. and Baker, B. (2015). African American Vernacular English and reading. In Lanehart, S., editor, *The Oxford Handbook of African American Language*.

Labov, W., Waletzky, J., and Weiner, J. (1993). The unobservability of structure and its linguistic consequences. In *Paper delivered at NWAV*.

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.8.3.

Levon, E. and Buchstaller, I. (2015). Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change*, 27(3):319–348.

Lewis, M. C. (2018). A critique of the principle of error correction as a theory of social change. *Language in Society*, 47(3):325–346.

Lippi-Green, R. (1997). *English with an accent*. Routledge, New York.

Litcofsky, K. A. and Van Hell, J. G. (2017). Switching direction affects switching costs: Behavioral, ERP and time-frequency analyses of intra-sentential codeswitching. *Neuropsychologia*, 97:112–139.

Lundquist, B. and Vangsnes, Ø. A. (2018). Language separation in bidialectal speakers: Evidence from eye tracking. *Frontiers in Psychology*, 9(AUG).

Maher, Z. K., Erskine, M. E., Byrd, A. S., Harring, J. R., and Edwards, J. R. (2021). African American English and early literacy: A comparison of approaches to quantifying nonmainstream dialect use. *Language, Speech, and Hearing Services in Schools*, 52(1):118–130.

Mallinson, C., Charity Hudley, A., Strickling, L. R., and Figa, M. (2011). A conceptual framework for promoting linguistic and educational change. *Linguistics and Language Compass*, 5(7):441–453.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190.

Molinaro, N., Kim, A., Vespignani, F., and Job, R. (2008). Anaphoric agreement violation: An ERP analysis of its interpretation. *Cognition*, 106(2):963–974.

Newkirk-Turner, B. L. and Green, L. (2016). Third person singular -s and event marking in child African American English . *Linguistic Variation*, 16(1):103–130.

Parker, D. (2019). Cue Combinatorics in Memory Retrieval for Anaphora. *Cognitive Science*, 43(3).

Patson, N. D. and Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, 69(5):950–971.

Preston, D. R. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Language Awareness*, 5(1):40–74.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roeper, T. and Green, L. (2008). Node labels and features: Stable and unstable dialects and variation in acquisition. *Linguistic Variation Yearbook*, 7:1–26.

Ryskin, R., Futrell, R., Kiran, S., and Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, 181(August):141–150.

Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., and Gibson, E. (2021). An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia*, 158(March):107855.

Salig, L. K., Kroff, J. R. V., Slevc, L. R., and Novick, J. M. (2021). Moving from bilingual traits to states: Understanding cognition and language processing through moment-to-moment variation. *Neurobiology of Language*, 2(4):487–512.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied linguistics*, 11(2):129–158.

Seymour, H. N., Roeper, T., and de Villiers, J. G. (2003). *DELV: Diagnostic evaluation of language variation: Screening test*. PsychCorp.

Spivey-Knowlton, M. and Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, 55(3):227–267.

Squires, L. (2014). Processing, Evaluation, Knowledge: Testing the Perception of English Subject-Verb Agreement Variation. *Journal of English Linguistics*, 42(2):144–172.

Squires, L. (2016). Processing grammatical differences: Perceiving versus noticing. In *Awareness and control in sociolinguistic research*, pages 80–103.

Squires, L. (2019). Genre and linguistic expectation shift: Evidence from pop song lyrics. *Language in Society*, 48(1):1–30.

Staum Casasanto, L. (2009). What do Listeners Know about Sociolinguistic Variation? *University of Pennsylvania Working Papers in Linguistics*, 15(2).

Steele, C. M. and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5):797.

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

Tamminga, M. and Wade, L. (2022). Coherence across social and temporal scales. In *The Coherence of Linguistic Communities*, pages 34–52. Routledge.

Terry, J. M., Hendrick, R., Evangelou, E., and Smith, R. L. (2010). Variable dialect switching among African American children: Inferences about working memory. *Lingua*, 120(10):2463–2475.

Terry, J. M., Hendrick, R., Evangelou, E., and Smith, R. L. (2015). Dialect Switching and Mathematical Reasoning Tests: Implications for Early Educational Achievement. In Lanehart, S., editor, *The Oxford Handbook of African American Language*.

Terry, J. M., Thomas, E. R., Jackson, S. C., and Id, M. H. (2022). African American English speaking 2nd graders , verbal –s , and educational achievement: Event related potential and math study findings. *PLoS ONE*, 17(10):1–29.

Valdés Kroff, J. R. and Dussias, P. E. (2023). Production, processing, and prediction in bilingual codeswitching. In *Psychology of Learning and Motivation*, pages 195–237. Elsevier.

Van Hofwegen, J. and Wolfram, W. (2017). On the Utility of Composite Indices in Longitudinal Language Study: The Case of African American Language. In Buchstaller, I. and Wagner, S. E., editors, *Panel studies of variation and change*, pages 73–95. Routledge, New York.

Vaughn, C. (2022). The role of internal constraints and stylistic congruence on a variant's social impact. *Language Variation and Change*, 34(3):331–354.

Vaughn, C. and Kendall, T. (2018). Listener sensitivity to probabilistic conditioning of sociolinguistic variables: The case of (ING). *Journal of Memory and Language*, 103(August):58–73.

Vaughn, C. R. (2019). Expectations about the source of a speaker's accent affect accent adaptation. *The Journal of the Acoustical Society of America*, 145(5):3218–3232.

Voeten, C. C. (2022). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. R package version 2.7.

Vorwerg, C. C., Suntharam, S., and Morand, M. A. (2019). Language control and lexical access in diglossic speech production: Evidence from variety switching in speakers of Swiss German. *Journal of Memory and Language*, 107(March):40–53.

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

Weissler, R. E. (2021). *Leveraging African American English Knowledge: Cognition and Multidialectal Processing*. Ph.d. dissertation, University of Michigan.

Weissler, R. E. and Brennan, J. R. (2020). How do Listeners Form Grammatical Expectations to African American Language? *U. Penn Working Papers in Linguistics*, 25(2).

Weldon, T. L. (2022). Visibly invisible: The study of middle class African American English. *Language and Linguistics Compass*, 16(11):e12477.

Wheeler, R. (2016). "So Much Research, So Little Change": Teaching Standard English in African American Classrooms. *Annual Review of Linguistics*, 2:367–390.

Wolfram, W. (1982). Language Knowledge and Other Dialects. *American Speech*, 57(1):3–18.

Wolfram, W. and Schilling, N. (2015). *American English: Dialects and Variation*, volume 25. Wiley.

Young, V. A. (2009). "Nah , We Straight": An Argument Against Code Switching. *JAC*, 29(1):49–76.

Zaharchuk, H. A., Shevlin, A., and Van Hell, J. G. (2021). Are our brains more prescriptive than our mouths? Experience with dialectal variation in syntax differentially impacts ERPs and behavior. *Brain and Language*, 218(January):104949.

Zhou, P., Garnsey, S., and Christianson, K. (2019). Is imagining a voice like listening to it? Evidence from ERPs. *Cognition*, 182(October 2018):227–241.