Spring 5-15-2022

The Effect of Social Primes on the Perception of Native- and Nonnative-Accented Speech

Drew J. McLaughlin
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological & Brain Sciences

Dissertation Examination Committee:
Kristin Van Engen, Chair
John Baugh
Calvin Lai
Jonathan Peelle
Mitch Sommers

The Effect of Social Primes on the Perception of Native- and Nonnative-Accented Speech

by

Drew J. McLaughlin

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2022
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

This work is supported by a National Science Foundation Doctoral Dissertation Research Improvement Grant awarded to Dr. Kristin J. Van Engen (PI) and Drew J. McLaughlin (co-PI), and a National Science Foundation Graduate Research Fellowship awarded to Drew J. McLaughlin.

To my PhD advisor, Kristin J. Van Engen: I am incredibly grateful to have had the opportunity to work with you for the last five years. Your support helped me thrive as a student, scientist, and human.

To my additional (unofficial) PhD advisor, Jonathan E. Peelle: Thank you for always encouraging me to take-on projects that expanded my interests and skillset. These projects have been some of my favorites during my time at Washington University in St. Louis.

To my first advisor, Melissa M. Baese-Berk: Thank you for encouraging me to pursue a PhD, even if it meant moving far from home. Pushing me "out of the nest" allowed me to establish my independence as a researcher, and I am very grateful.

To my parents, Katherine C. and Douglas E. McLaughlin: Thank you for the support you have shown me in my academic endeavors all my life. I could not have reached this point without you.

To my partner, Ryland L. Miller: Thank you for putting up with the Missouri humidity for four years so that I could pursue my dreams. Your love, patience, and support mean the world to me.

Drew J. McLaughlin

Washington University in St. Louis

May 2022

ABSTRACT OF THE DISSERTATION

The Effect of Social Primes on the Perception of Native- and Nonnative-Accented Speech

by

Drew J. McLaughlin

Doctor of Philosophy in Psychological and Brain Sciences

Washington University in St. Louis, 2022

Assistant Professor Kristin J. Van Engen, Chair

Listeners use more than just acoustic information when processing speech. Social information, such as a speaker's race/ethnicity, can also affect listeners' understanding of the speech signal. In some cases, these *social primes* can facilitate perception, while in others they may inhibit perception. Indeed, a picture of an East Asian face has been shown to facilitate the perception of Mandarin Chinese-accented English but interfere with the perception of American-accented English. The present dissertation builds on this line of inquiry, addressing novel topics including the generalizability and specificity of social priming effects, their relationship with implicit racial/ethnic associations, and their role in perceptual adaptation to nonnative-accented speech. In four online experiments, we examined the effect of visual race/ethnicity guises on transcription accuracy for native-accented and nonnative-accented speech (presented in background noise). Results of these experiments were mixed. Our first experiment successfully replicated prior work, demonstrating that an East Asian prime can facilitate perception of Mandarin-accented English speech. However, despite our attempt to assure sufficient power via large sample sizes, in all subsequent experiments we did not find effects of social primes on

speech transcription accuracy or perceptual adaptation. Notably, our null outcomes have one positive implication: Minority race/ethnicity speakers with native-accented English speech may not be at a perceptual disadvantage as compared to White speakers (as indicated by prior work). Ultimately, further examination of social priming is needed to determine whether the mixed findings of the current work may reflect a small or a context-dependent effect. We suggest that by including in-depth analyses of subjects' social networks (i.e., to determine their exposure to racial/ethnic diversity and relevant accent varieties), future investigations will be able to assess individual differences in susceptibility to social priming effects.

1. Introduction

The speech signal conveys more than just a message. Paralinguistic information, such as cues as to the speaker's race, gender, and social class, is also embedded within the speech stream. Listeners can exploit these indexical cues to identify social characteristics of the speaker (Kraus, Torrez, Park, & Ghayebi, 2019), as well as to parse the speech stream (Ladefoged & Braoadbent, 1957; Johnson, Strand, & D'Imperio, 1999). Particularly notable is the effect that social cues from outside the acoustic signal can have on speech processing. For example, perceived talker gender (manipulated with video) can influence the categorization of fricatives in an /s/ to /ʃ/ continuum (Strand, 1999), and perceived speaker race (manipulated with still images) can affect the accuracy of overall speech recognition (Babel & Russell, 2015; McGowan, 2015). In other words, listeners can be *socially primed* to expect a specific socio-phonetic pattern from a given speaker, and this can either facilitate or interfere in the speech perception process.

This dissertation will investigate the effect of speaker race/ethnicity information on the perception of native- and nonnative-accented speech for White American listeners. In a series of experiments, we will integrate group-wide behavioral data and individual differences data to examine the following: 1) The generalizability and specificity of social priming effects; 2) Whether there is a link between implicit associations and the effects of social primes on speech perception; and, 3) Whether social primes can benefit the listener during perceptual adaptation to unfamiliar, nonnative-accented speech.

1.1 Background

1.1.1 Nonnative Accents, Recognition Accuracy, and Adaptation

Nonnative accents are characterized by their systematic segmental and suprasegmental deviations from native accent norms. For listeners, these deviations can lead to mismatches between external acoustic information and internal perceptual categories during speech processing (Van Engen & Peelle, 2014), and may ultimately result in poorer speech recognition accuracy (or "intelligibility"). Here, and in prior work (McLaughlin, Baese-Berk, Bent, Borrie, & Van Engen, 2018), we operationalize recognition accuracy as the proportion of keywords correctly transcribed from a target sentence. Listeners often have poorer recognition accuracy for unfamiliar nonnative-accented speech than native speech (though this depends on speaker fluency; Munro & Derwing, 1995), but individual listener differences, such as working memory capacity (McLaughlin et al., 2018) and vocabulary knowledge (Banks, Gowen, Munro, & Adank, 2015; Bent, Holt, Miller, & Libersky, 2019; McLaughlin et al., 2018), can also impact listening performance.

Additionally, while it may be difficult to process a nonnative accent initially, listeners can rapidly adapt to (or "accommodate") unfamiliar nonnative-accented speech (Clarke & Garrett, 2004), thereby reducing the cognitive demands of speech processing (Brown, McLaughlin, Strand, & Van Engen, 2020). By tracking perceptual processing speed with a reaction time-based word identification task, Clarke and Garrett (2004) demonstrated rapid adaptation by native English listeners to both a Spanish- and a Mandarin Chinese-accented speaker of English. The authors also compared improvement following training (i.e., exposure to stimuli from a given condition) with nonnative-accented sentences versus native-accented sentences presented in background noise. Although all subjects showed improved reaction times across blocks, in a post-test session with nonnative accent, subjects trained with nonnative accent outperformed subjects trained with native speech presented in noise. This outcome demonstrated that perceptual accommodation of a nonnative accent may rely on a separate mechanism than accommodation of speech in noise, and follow a framework of transfer appropriate processing (see Lockhart, 2002). The results of McLaughlin et al. (2018) tell a similar story: correlations in performance across multiple adverse listening conditions (native English speech in speech-shaped noise, native English speech in multi-talker babble, nonnative English speech in quiet, and nonnative English speech in speech-shaped noise) indicated that monolingual English subjects' performance when listening to nonnative English speech in quiet was unrelated to their performance when listening to native English speech in either noise condition. In other words, listeners who are good at understanding and accommodating nonnative accent are not necessarily good at understanding speech presented in noise. Altogether, these studies indicate that perceptual adaptation to nonnative accent is a uniquely challenging process.

Notably, perceptual adaptation occurs for a single talker as well as multiple talkers, and perceptual learning can generalize to a novel talker of the same accent. A seminal study by Bradlow and Bent (2008) found in native English listeners that training with multiple nonnative speakers of the same accent resulted in better recognition accuracy at post-test of the training talkers and of novel talkers with the same accent (as compared to training with a single nonnative talker). Further, their results indicated that multi-talker accent training may be more effective than single-talker accent training, benefiting the listener when they encounter a novel nonnative accent (e.g., encountering a Slovakian accent after training with a Mandarin accent). Work by Baese-Berk, Bradlow, and Wright (2013) with native English listeners expanded on this latter finding, further demonstrating that training with multiple nonnative talkers of *different* language backgrounds can also improve performance on a post-test with a novel talker with an unfamiliar/untrained nonnative accent as compared to training with a single talker (see also Bieber & Gordon-Salant, 2017).

Altogether, work examining accommodation of nonnative accent indicates that listeners not only adapt to a given nonnative speaker, but that they can learn the systematic deviations present across multiple speakers with nonnative accent and generalize that learning. This evidence supports an interpretation of accent accommodation as a process in which listeners "tune" their own perceptual categories to more easily map to the productions of a nonnative talker (referred to as *recalibration*). Indeed, work examining acoustically-manipulated native accent also provides strong evidence that listeners can adjust their phonemic category boundaries (Norris, McQueen, & Cutler, 2003), although efforts to directly link this recalibration mechanism to nonnative accent perception are mixed (Xie, Theodore, & Myers, 2017; Zheng & Samuel, 2020).

What remains unclear is whether there may be multiple perceptual mechanisms supporting processes of speaker, accent, and speech-in-noise accommodation, and whether a given mechanism may only be utilized under specific circumstances. For example, the multi-accent accommodation documented by Baese-Berk et al. (2013) is characteristically different from multi-speaker (same accent) accommodation: Listeners are exposed to multiple nonnative accents that each differ in how they systematically deviate from native accent productions. A recalibration account of accent accommodation cannot explain this outcome alone. Thus, an additional accent accommodation mechanism that has been proposed is *criteria relaxation*. As discussed by Zheng and Samuel (2020), it may be the case that when faced with accent variation listeners relax their thresholds for accepting input as a particular phoneme, lexical item, and so on. Relatedly, listeners may rely less on categorical boundaries when processing nonnative accent; they may encode phonetic variation within-category, registering acoustic details that would typically (with an native speaker) be ignored.

However, little empirical evidence currently exists to support the conclusion that criteria relaxation supports accent accommodation. Maye, Aslin, and Tanenhaus (2008) examined American English listeners' adaptation to an regional dialect of American English, created by artificially shifting the formants of vowels. In their study, they only found evidence of recalibration (measured via shifted category boundaries), and not criteria relaxation. However, this same question has yet to be examined with nonnative accent perception. Addressing whether nonnative accent accommodation is supported by recalibration and/or criteria relaxation mechanisms is not one of the aims of the present dissertation, however we anticipated that our examination of the effects of social information on perceptual adaptation may provide valuable insights on this topic.

1.1.2 Social Priming

While the proposed study will focus on the effects of visual race/ethnicity guises on speech recognition, many notable social priming studies have examined the effects of verbal guises on speech recognition. In a sample of subjects born and raised in Detroit, Niedzielski (1999) found that information about a speaker's nationality (i.e., a speaker guise of "from Detroit, USA" or "from Ontario, Canada") changed the perception of vowels, even though the same stimuli (produced by a Detroit speaker) were presented to each group. This same effect has also been found in New Zealanders, who reported hearing more Australian-like vowel pronunciations when primed with "Australian" than when primed with "New Zealand" (Hay, Nolan, & Drager, 2006; see also McGowan & Babel, 2019).

Only one study has examined the effect of social information on perceptual adaptation to nonnative accent. Vaughn (2019) compared native English subjects' adaptation to the same nonnative Spanish-accented speaker of English following one of three verbal guises: no guise (no information provided), native-accent guise (e.g., 'the speaker's first language is English, but his parents are Spanish speakers'), and nonnative accent guise (e.g., 'the speaker's first language is Spanish'). Similar to what would be expected based on previous manipulations of race information (e.g., McGowan, 2015), providing language background information about the speaker led to improved recognition accuracy, and facilitated adaptation to the accent. Surprisingly, however, subjects given the native-accent guise actually outperformed subjects given the nonnative accent guise (despite the former being accurate). This outcome indicated that ethnicity information (i.e., implicating that the speaker was Latino) may have facilitated perception of the Spanish accent, but that listener biases toward nonnative speakers of English may have also interfered with perception.

Visual guises can also negatively affect how well listeners comprehend speech. Rubin (1992) found that comprehension of a short lecture (as measured with a cloze test) was poorer for American subjects who listened while viewing a picture of an East Asian face than when viewing a White face, despite the speech in both conditions being from the same native American-accented speaker (see also Kang & Rubin, 2009). Babel and Russell (2015) and McGowan (2015) are the only studies that have examined the direct effect of social priming on speech recognition accuracy for native English listeners. Babel and Russell found that Canadian listeners' recognition accuracy for native Canadian-accented English speech presented in babble was reduced for Chinese-Canadian talkers when still images of the talkers were shown on screen (as compared to presenting the audio with a fixation cross only). For the White-Canadian talkers presented in the same experiment this was not the case, indicating that biases against the Chinese-Canadian talkers may have affected speech perception. In contrast, however, McGowan (2015) found that recognition accuracy of Mandarin Chinese-accented speech presented in babble was *better* for American listeners when subjects were presented with an East Asian prime, as compared to a White prime. Together, these results indicate that the outcome of Babel and Russell (2015) may actually reflect an automatic social priming cost. In other words, the faces of the Chinese-Canadian speakers presented in Babel and Russell's experiment may have activated socio-phonetic categories for nonnative, Chinese-accented English; thus, when listeners encountered native, Canadian-accented English speech, this incongruency may have hindered speech recognition accuracy. Complementary to this finding, the results of McGowan (2015) demonstrate that Asian faces do not negatively affect speech perception under all circumstances, but can actually facilitate perception of a Mandarin Chinese accent.

This interpretation of prior work is in line with exemplar models of speech perception (Hintzman, 1984; Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001). Unlike a normalization model of speech perception, in which variation related to social characteristics of the speaker would be stripped away, exemplar theory proposes that phonetically-detailed episodic traces (or "exemplars") are stored in the lexicon. On such a view, the incoming acoustic signal activates acoustically-similar exemplars, which assists with recognition of the acoustic signal. The efficiency of recognition is influenced by the number of similar exemplars in memory, meaning that speech processing for more familiar speakers (and accents) should be faster. Indeed, a wealth of research has indicated that listeners encode talker-specific voice attributes ("surface details"; Goldinger, 1996) and that talker familiarity affects the efficiency of speech perception (Newman & Evers, 2007; Magnuson, Nusbaum, Akahane-Yamada, & Saltzman, 2020). For example, word recognition is faster and more accurate when target words are presented in the same voice at exposure and test (Craik & Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993).

Under exemplar theory, the phonetic patterns of exemplars are also linked to social information and can be activated by social characteristics of the speaker. In other words, non-linguistic information, such as a speaker's perceived gender (e.g., Strand, 1999) or race/ethnicity (e.g., McGowan, 2015), would be able to activate linguistic exemplars and facilitate perception of speech. Thus, based on an exemplar account we would also expect that activation of socio-phonetic categories ("social priming") should vary across listeners based on their unique experiences. If this is the case, we would expect that individuals with stronger associations between given racial/ethnic and accent categories would have larger social priming effects.

It is important to note that, in addition to exemplar theory, other models and theories of speech perception may also provide a framework for characterizing and explaining the social priming phenomenon. In particular, the *ideal adaptor model* (Kleinschmidt & Jaeger, 2015; Kleinschmidt, Weatherholtz, & Jaeger, 2018) is a probabilistic learning model that can accurately capture the process of perceptual adaptation to accent as well as socio-linguistic inferences such as social priming effects. Similar to exemplar theory, the ideal adaptor model approaches the "problem" of lack of invariance (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) in speech perception differently: By tracking *lawful variability* (Elman & McClelland, 1983), listeners can overcome variability in the speech signal, and even leverage socio-indexical cues to facilitate speech perception. As is the case for exemplar theory, the ideal adapter model posits that the listener's experience should affect social priming. In other words, some listeners will have greater knowledge of the probabilistic co-occurrences of social cues (such as perceived race or ethnicity of the speaker) and linguistic cues (such as accent qualities) than others, and these differences in listeners' prior knowledge should affect how social primes affect their perception of native- and nonnative-accented speech.

One of the aims of the present dissertation is to examine the hypothesis that individuals with stronger associations between given racial/ethnic and accent categories will have larger social priming effects. By examining this hypothesis, we aimed to improve our understanding of the underlying mechanism supporting social priming effects, and better situate social priming effects within current models and theories of speech perception (such as, but not limited to, exemplar theory and the ideal adaptor model). To this end, we incorporate individual difference measures of implicit associations and language attitudes alongside our social priming experiment in Chapter 3.

1.1.3 Implicit Associations

Self-reported attitudes provide valuable insight into stigmatization and stereotyping of social groups, but may also represent suppressed versions of subjects' actual attitudes (Wilson & Dunn, 2004); for example, for attitudes toward nonnative accents subjects wishing to maintain a non-prejudiced outward appearance may avoid expressing negative evaluations of speakers. By measuring subjects' *implicit* racial/ethnic associations in tandem with their explicit attitudes, we can ascertain whether these internalized and externalized attitudes deviate.

Increasingly common in psychological research, implicit measures are useful for revealing individual differences in underlying associations and biases. The most common tool for examining implicit associations is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT is a matching task that measures automatic associations between two contrasted constructs (for a review, see Schnabel, Asendorpf, & Greenwald, 2008). In the IAT, different constructs are grouped together in each block. For example, in one block, subjects may be told to sort images of White faces and American places into the same category, and images of East Asian faces and foreign places into a different category; then, in a different block these pairings would be reversed (i.e., White faces with foreign places, East Asian faces with American places). The key assumption of the IAT is that sorting related constructs into the same category will be easier and faster. By comparing response times across blocks, the IAT can thus measure the strength of associations between these sets of contrasted constructs, and allow the researcher to draw conclusions about subjects' implicit associations and biases.

Numerous types of IATs exist for examining implicit associations. Pictures of faces are often employed in IATs (e.g., for examining implicit biases related to race and age), but orthographic stimuli can also be used for constructs that cannot be captured visually (e.g., for the constructs good and bad). A seminal example of the IAT is work by Devos and Banaji (2005), in which the authors investigated implicit biases of White, Asian, and African Americans. Across six experiments, Devos and Banaji examined associations between racial groups (using images of faces) and the constructs *American* vs. *Foreign* (using images of iconic American and non-American scenes). Overall, their data indicated a general bias toward associating White American faces with the construct American more than Asian American or African American faces. When examining each racial group of subjects separately, however, the authors found that Asian American subjects themselves show a stronger association between the category American and White faces than between the category American and Asian faces, but that Black American subjects showed equal associations for both Black and White faces with the construct American. The results of Devos and Banaji's work demonstrated that implicit biases held by Americans can be internalized by racial minority groups, but that this phenomenon does not occur for all racial minority groups in the United States.

Implicit association tests have also been used in linguistic contexts. Using linguistic IATs (both orthographic and auditory), Campbell-Kibler (2012) found that the word-final American Southern variant *-in* ([ɪn] or [ən]) and American Northern variant *-ing* ([ɪŋ]) were implicitly associated with blue collar professions (e.g., plumber) and white collar professions (e.g., doctor), respectively, for American subjects. Additionally, Pantos and Perkins (2013) found with an auditory IAT that American listeners more strongly associate American-accented English with the construct good and Korean-accented English with the construct bad. Together, these studies indicate a connection between accentedness and implicit biases towards social groups defined by geographic region and race/ethnicity.

The relationship between implicit racial associations and social priming effects was initially explored in Babel and Russell (2015). As discussed in Section 1.1.2, Babel and Russell found that Canadian listeners had poorer recognition accuracy for native-accented Canadian English speech when a picture of the Chinese Canadian speaker's face was presented onscreen than when a fixation cross was presented onscreen (indicating a negative social priming effect). As part of this same study, the authors also examined listeners' implicit racial biases using an orthographic IAT that measured associations between common White and Chinese Canadian surnames and the constructs Positive vs. Negative. The overall group outcome from the IAT indicated that subjects had stronger associations between White surnames and the construct Positive, and between Chinese surnames and the construct Negative, than the opposite combination. Babel and Russell attempted a correlation between individual subjects' IAT scores and a summarized measure of the social priming cost observed in the speech perception task, and did not find a significant relationship.

The only study that has found a relationship between IAT scores and performance on a speech perception task was work by Yi, Phelps, Smiljanic, and Chandrasekaran (2013). In their study, Yi and colleagues found that monolingual American English listeners derived less benefit from adding the video signal to the corresponding audio signal ("audiovisual benefit") for Korean-accented English talkers than American-accented English talkers. Subjects in the study also completed an IAT with White vs. Asian faces and the constructs American vs. Foreign (similar to Devos & Banaji, 2005). Group-wide, the IAT revealed stronger implicit associations between the White faces and the construct American, and between Asian faces and the construct Foreign, than the opposite combinations. Additionally, a significant correlation emerged between individual subjects' IAT scores and a summary statistic of their reduced audiovisual benefit for Korean-accented as compared to American-accented English. Yi and colleagues interpreted this correlation as an indicator that biases toward Asian speakers may negatively affect the process of audiovisual integration for speech. However, in a direct replication of this study with a larger sample size ($N = 260$ as compared to $N = 19$ in Yi et al., 2013) McLaughlin, Brown, Carraturo, and Van Engen (2022) did not find evidence that IAT scores were related to reduced audiovisual benefit for Korean-accented English. The main difference in audiovisual benefit for Korean-accented versus American-accented speakers successfully replicated, and a follow-up experiment further demonstrated that this finding was not due to a confound of the overall difficulty level of each accent type. In summary, audiovisual benefit appears to be reduced for nonnative accent as compared to native accent, regardless of the overall intelligibility level of speech, but this difference in audiovisual benefit does not appear to be caused by listeners' biases against the Korean talkers. As would be expected based on the Principle of Inverse Effectiveness, the degree of audiovisual benefit was related to the overall difficulty of speech perception (i.e., when there was more room to improve, subjects showed more benefit), but there was an additional cost to audiovisual integration for nonnative accent as compared to native accent.

The findings of McLaughlin and colleagues (2022) suggest that the initial correlation found by Yi and colleagues (2013) between IAT scores and audiovisual integration may have been a spurious outcome due a small sample size. Indeed, although IATs are widely used and validated as measures of implicit associations, using them as measures of individual differences can prove challenging and requires a substantial sample size to ensure sufficient power. The internal reliability of IATs varies by construct (for example, IATs examining political preferences tend to have higher reliability than IATs examining racial attitudes; Greenwald, Poehlman, Ulmann, & Banaji, 2009), and even correlations with behavioral measures of explicit attitudes can be very small (e.g., $r = .24$ for race IATs; Greenwald et al., 2009). Thus, it is unclear what size of correlation one would expect between an IAT and a speech perception measure, but reasonable to expect that it would be small. In McLaughlin et al. (2022), the authors estimated a sample size based on a power analysis for a correlation of $r = .20$ (a conservative estimate more than half the size of the correlation originally found by Yi et al., 2013). For our examination of social priming effects in the present dissertation, we also used a conservative power analysis to estimate a sample size that would ensure sufficient power to examine relationships between IAT scores and social priming (discussed further in Chapter 3).

1.1.4 Language and Social Attitudes

Not all accents are viewed equally among language users, and, within a geographic region, there is usually one accent (or one set of similar accents) collectively regarded as the "standard" (Lippi-Green, 2012). For example, in the United States, the standard American accent is a derivative of a variety spoken in the Midwest by primarily White speakers, and in the United

Kingdom, the standard British accent (also referred to as Received Pronunciation and the Queen's English) is a derivative of a variety spoken in Southern England. In both cases, the accents referred to as the standard are not linguistically superior forms of English; rather, they are accents originally used by groups of speakers with higher social status and more influence (Lippi-Green, 2012).

Speakers with non-standard accents often face prejudice and stigmatization that impacts their everyday lives (Carlson & McHenry, 2006; Purnell, Idsardi, & Baugh, 1999). For nonnative-accented speakers in particular, the comprehensibility of their speech affects how listeners perceive their intelligence (Bresnahan, Ohashi, Nebashi, Liu, & Morinaga Shearman, 2002), and credibility (De Meo, 2012). Additionally, while both nonnative and regional speakers report similar experiences of stigmatization, speakers with nonnative accents report more communication difficulties and a lesser sense of belonging within the United States (Gluszek & Dovidio, 2010).

A large amount of research examining accent perception has used numeric rating scales (following presentations of voice recordings) to measure listener attitudes toward speakers. Numerous traits have been assessed by researchers, and in many cases these map onto two distinct dimensions: *status* and *solidarity* (for factor analyses, see Brennan & Brennan, 1981; Dragojevic & Giles, 2014; Heaton & Nygaard, 2011). The status dimension typically captures traits such as intelligence, confidence, success, ambition, class, and education, while the solidarity dimension captures traits such as trustworthiness, pleasantness, sincerity, kindness, friendliness, and sociality. In some cases, a third dimension called *dynamism* (e.g., liveliness) has also been identified (see Fuertes, Gottdiener, Martin, Gilbert, & Giles, 2012). In the social psychology literature, the status and solidarity dimensions would be better recognized as *competence* and *warmth*, as in the Stereotype Content Model (Fiske, Cuddy, Glick, and Xu, 2002).

Listeners' attitudes are typically influenced by speech comprehensibility. For example, more comprehensible nonnative-accented speakers tend to be rated more positively on status traits than less comprehensible nonnative-accented speakers (Bresnahan et al., 2002). What's more, Dragojevic and Giles (2016) found for English listeners that even when the same nonnative speaker was presented, but either in quiet or in white noise (the latter of which was less comprehensible), the nonnative accent in noise was rated lower for status traits and subjects also reported poorer affect (i.e., poorer mood). In two experiments examining perception of Punjabi- and then Mandarin Chinese-accented English by English listeners, Dragojevic et al. (2017) tested whether the relationship between nonnative accent strength and speaker evaluations (status and solidarity) was mediated by speaker comprehensibility, listeners' affective response (i.e., mood), and/or speaker prototypicality. In a mediation path analysis, comprehensibility and the sequential effect of comprehensibility on negative affect mediated the relationship between nonnative accent strength and status evaluations. Prototypicality judgments, on the other hand, did not mediate this relationship, indicating that negative status evaluations are largely based on the comprehensibility of speech.

There are real-world impacts of the stigmatization of non-standard accents. Most listeners can identify a speaker's gender, age, race, and social class from brief, out-of-context speech samples alone (Kraus et al., 2019), and speakers with non-standard accents can have greater difficulty securing employment (Carlson & McHenry, 2006) and housing (Purnell et al., 1999) opportunities. These self-reported attitudes provide valuable insight into the stigmatization of accented speech, but may also represent suppressed versions of subjects' actual attitudes (Wilson

& Dunn, 2004). By measuring subjects' implicit racial/ethnic associations in tandem with their explicit attitudes, we aimed to ascertain whether these internalized and externalized attitudes deviate, and how each may relate to social priming effects.

Generalizability of social priming. One of the aims of the present dissertation is to determine whether social priming effects occur for race and ethnicity primes other than White and East Asian. If the source of social priming effects is socio-phonetic categories linking race/ethnicity information with accent qualities, then we predict that American listeners may expect nonnative accents for some additional races and ethnicities. As shown in **Figure 1.1**, Zou and Cheryan's (2017) racial position model of the United States indicates that some racial/ethnic groups, including Asian Americans, Latinx Americans, and Middle Eastern Americans, are more strongly associated with *cultural foreignness* than Black and White Americans; here, cultural foreignness includes multiple stereotypes regarding foreign (nonnative) accent as well as some related to immigration and being "outsiders". Notably, Asian, Latinx, and Middle Eastern Americans (as well as Black and White Americans) are also distinguished on a separate status scale, which captures stereotypes related to wealth, education, and power. Based on this model of Americans' racial and ethnic stereotypes, we predicted that social priming effects would occur for Latinx and Middle Eastern primes in a similar manner as they do for East Asian primes. On this view, one would also expect that Black primes may result in similar outcomes as White primes – although this prediction is not tested in the present dissertation.

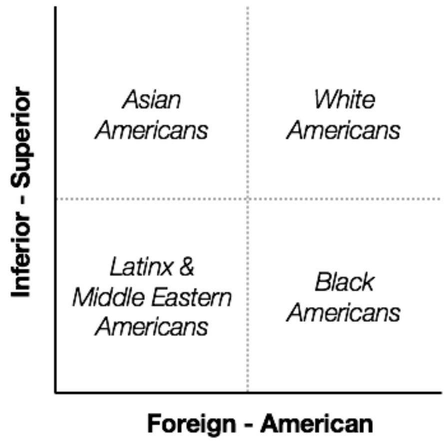

Figure 1.1. Reproduction of Zou & Cheryan's (2017) racial position model of the United States. The cultural foreignness scale on the x-axis separates race/ethnicity groups by how strongly they are associated with foreign accents (among other traits). The y-axis is a measure of status, which also distinguishes racial groups by stereotypical wealth, education, and power.

Does *Foreign* = Foreign accent? As discussed above, in Devos & Banaji's (2005) seminal work the authors find a general bias for White and Asian (but not African) Americans toward associating White American faces with the construct *American* more so than Asian American or African American faces, and Asian and African faces with the construct *Foreign* more so than White American faces. These findings spurred a wealth of research examining Americans' implicit associations, as well as their explicit attitudes and stereotypes, of the "Americanness" of various racial and ethnic groups in the United States. However, it is unclear to what degree attitudes toward "foreign" (nonnative) accents are intertwined with the general construct of foreignness.

Indeed, based on Zou & Cheryan's (2017) racial position model, we might predict that Black Americans would not be implicitly associated with the construct foreign (as found by Devos & Banaji, 2005), because based on measures of Americans' explicit attitudes they appear to be aligned with White Americans on a *cultural foreignness* scale (**Figure 1.1**). This may indicate that implicit measures such as an *American* vs. *Foreign* IAT capture ingroup vs. outgroup associations more broadly. For example, for White Americans, all non-White American faces may be less strongly associated with the construct "American" and more with the construct "Foreign". If this is the case, then an *American* vs. *Foreign* IAT may be a less robust measure of implicit associations between race/ethnicity and expectations about foreign accentedness. However, the *American* vs. *Foreign* IAT is much more established in the literature (and used in linguistic research such as Yi et al., 2013 and McLaughlin et al., 2022), and we determined it to be the most suitable measurement tool for examining individual differences in the present dissertation. With this potential limitation in mind, we decided to include a second IAT of *Good* vs. *Bad* associations that could provide dissociative evidence. As will be discussed in Chapter 3, in the present dissertation we predicted that individual differences in the strength of implicit associations between racial/ethnic categories and the constructs *American* vs. *Foreign* would predict social priming in a speech perception task; in contrast, we predicted that individual differences in the strength of implicit associations between racial/ethnic categories and the constructs *Good* vs. *Bad* (as measured with a different IAT) would *not* predict social priming in a speech perception task.

1.1.5 Summary

The evidence discussed thus far provides a compelling case for an effect of social information on speech perception. However, many important questions remain unaddressed by prior work. For example, these effects have yet to be examined outside of the context of White and East Asian primes and native- and Mandarin Chinese-accented speech. Additionally, it is unclear whether these effects on speech perception may be related to individual differences in listeners' implicit racial/ethnic associations (e.g., how strongly they associate a given race/ethnicity with the constructs American vs. Foreign). The current work will address these gaps in the literature and extend the examination of social priming effects to novel areas, such as perceptual adaptation to nonnative accent.

1.2 Objectives and Hypotheses

Here, we outline the objectives and hypotheses of the present dissertation. Notably, in this chapter's reviewed literature, participants in all studies were typically representative, convenience samples (i.e., non-restrictive based on subject race or ethnicity) from college populations. In many cases, summaries of subjects' reported races and ethnicities are not provided by the authors. In Experiment 1 of the present dissertation, data from a representative, convenience sample of American college subjects is also collected, and we explore the effect of listener race on social priming effects. Based on this exploratory analysis, we opted to focus on White American listeners for all further experiments, anticipating that priming effects would be most robust in this population. Thus, our objectives, hypotheses, and conclusions focus on social priming effects for White American subjects.

Objective 1: We aim to test the *generalizability* of social priming effects by examining them in the context of novel races and accents. Specifically, we will examine White Americans' perception of native-accented speech following Latinx vs. White primes (**Chapter 3: Experiment 3**) and the perception of nonnative, Arabic-accented speech following Middle Eastern vs. White primes (**Chapter 2: Experiment 2**).

Hypothesis 1: Based on Zou and Cheryan's (2017) racial position model of the United States (**Figure 1.1**), we predict that subjects will have poorer transcription accuracy for native

(American) accented English speech when primed with a Latinx face than when primed with a White face (as has been found for East Asian faces; Babel & Russell, 2015).

Hypothesis 2: We predict that listeners will have better recognition accuracy for Arabic-accented English speech when viewing a Middle Eastern face than when viewing a control or White face, as has been found for Mandarin-accented English speech when primed with an East Asian face as compared to a control or White face.

Objective 2: We aim to test the *specificity* of social priming effects for nonnative speech (**Chapter 1, Experiment 2**), and propose two competing hypotheses.

Hypothesis 1A: Social primes have a high degree of specificity, such that White American listeners associate specific races/ethnicities with specific accents. Thus, for Mandarin-accented English, viewing a Middle Eastern prime (i.e., a minority race/ethnicity prime that is uncommon for the given nonnative accent) will incur a cost to listening performance as compared to a control or an East Asian prime. Similarly for Arabic-accented English, viewing an East Asian prime will incur a cost to listening performance as compared to a control or a Middle Eastern prime. These outcomes would indicate that minority race/ethnicity primes that are perceived as 'incongruent' with a given nonnative accent impose a similar cost effect as has been found for White (majority race) primes for Mandarin-accented English (McGowan, 2015).

Hypothesis 1B: Social primes are not highly specific; rather, White American listeners broadly associate the majority race (White) with a native accent and minority races/ethnicities with nonnative accents. Thus, viewing a minority race/ethnicity prime that is uncommon for a given nonnative accent (e.g., a Middle Eastern face with a Mandarin

Chinese accent) will not incur a cost to listening performance as compared to a control or "congruent" prime (an East Asian face with a Mandarin Chinese accent) – and may even provide a benefit. For example, it may be the case that when primed with a Middle Eastern face, subjects are not activating relevant exemplars (or "*recalibrating"* their phonemic categories) for an Arabic accent, but rather broadly preparing for a nonnative accent of any type. This prediction, in particular, would be in line with a *criteria relaxation* hypothesis: Listeners may relax their thresholds for accepting input as a particular phoneme or lexical item when anticipating difficult accent variation; relatedly, listeners may rely less on categorical boundaries when anticipating a nonnative accent, encoding phonetic variation within-category and registering acoustic details that would typically (with a native speaker) be ignored. In this way, even a minority race/ethnicity prime that is less typical for a given accent would have a different effect than a White (majority) race prime, and could provide a benefit to the listener.

Objective 3: We aim to test whether there is a link between implicit associations and the effects of social primes on speech perception by examining individual differences (**Chapter 3: Experiment 3**).

Hypothesis: We predict that individual listeners' implicit associations will be related to the size of their "social priming costs". That is, for a given listener, the size of the difference between conditions (e.g., performance of White Americans following a White versus an East Asian prime on native vs. Mandarin-accented English speech) may be larger or smaller, and we expect that listeners with larger priming effects will have stronger implicit associations between these races/ethnicities and the constructs *American* vs. *Foreign*.

Objective 4: We aim to investigate whether social primes can benefit White American listeners during perceptual adaptation to unfamiliar, nonnative-accented speech. Specifically, we will examine single-talker adaptation (**Chapter 2: Experiments 1 and 2**) and speaker-independent (generalizable) adaptation (**Chapter 4: Experiment 4**).

Hypothesis 1: For single-talker adaptation, we predict an immediate benefit of 'congruent' primes over a control prime and over a White prime. Specifically, for a Mandarin Chinese accent we expect White American subjects to perform better at the onset of the task when shown an East Asian face (as compared to the control or White face), and for a Arabic accent we expect White American subjects to perform better at the onset of the task when shown a Middle Eastern face (as compared to the control or White face). Additionally, we predict that the White prime will impose a prolonged negative effect on perceptual adaptation for both accent conditions; in other words, as compared to the control prime (which has no race/ethnicity information), the rate of improvement across the trials in the task will be hindered for the White prime condition.

Hypothesis 2: For speaker-independent (generalizable) adaptation, we predict that social primes will impact the effectiveness of nonnative accent training. Specifically, we predict that White American subjects assigned to a multi-talker Cantonese accent training in which East Asian primes co-occur with the auditory stimuli will perform better during training and at post-test than White American subjects assigned to a version of this training in which White primes co-occur with the auditory stimuli.

2. Experiments 1 and 2: Social Priming Effects for Nonnative Speech

In this study, we began by replicating the social priming effect found in McGowan (2015) with a larger sample of subjects (Experiment 1). Next, we expanded on those findings by investigating another nonnative accent and additional priming conditions (Experiment 2). Our aims were to test both the *generalizability* and the *specificity* of social priming effects for nonnative speech. To test the *generalizability* of social priming effects, we mirrored the design of McGowan (2015) with an Arabic accent and Middle Eastern prime (vs. a control and White prime). We predicted that listeners would have better recognition accuracy for Arabic-accented English when viewing a Middle Eastern face than when viewing a control or White face (as was the case for a Mandarin accent and East Asian prime). To test the *specificity* of social priming effects, we created two conditions with unexpected combinations of a minority race/ethnicity prime and a nonnative accent. We posited two competing hypotheses:

1) Social primes have a high degree of specificity, such that White American listeners associate specific races/ethnicities with specific accents. Thus, for Mandarin-accented English, viewing a Middle Eastern prime (i.e., a minority race/ethnicity prime that is uncommon for the given nonnative accent) will incur a cost to listening performance as compared to a control or an East Asian prime. Similarly for Arabic-accented English, viewing an East Asian prime will incur a cost to listening performance as compared to a control or a Middle Eastern prime. These outcomes would indicate that minority race/ethnicity primes that are perceived as ‘incongruent’ with a given nonnative accent

impose a similar cost effect as has been found for White (majority race) primes for Mandarin-accented English (McGowan, 2015).

2) Social primes are not highly specific; rather, White American listeners broadly associate the majority race (White) with a native accent and minority races/ethnicities with nonnative accents. Thus, viewing a minority race/ethnicity prime that is uncommon for a given nonnative accent (e.g., a Middle Eastern face with a Mandarin Chinese accent) will not incur a cost to listening performance as compared to a control or "congruent" prime (an East Asian face with a Mandarin Chinese accent) – and may even provide a benefit. For example, it may be the case that when primed with a Middle Eastern face, subjects are not activating relevant exemplars (or "*recalibrating"* their phonemic categories) for an Arabic accent, but rather broadly preparing for a nonnative accent of any type. This prediction, in particular, would be in line with a *criteria relaxation* hypothesis: Listeners may relax their thresholds for accepting input as a particular phoneme or lexical item when anticipating difficult accent variation; relatedly, listeners may rely less on categorical boundaries when anticipating a nonnative accent, encoding phonetic variation within-category and registering acoustic details that would typically (with an native speaker) be ignored. In this way, even a minority race/ethnicity prime that is less typical for a given accent would have a different effect than a White (majority) race prime, and could provide a benefit to the listener.

Notably, Hypothesis 1 is in line with an exemplar theoretic account of social priming. If listeners are storing phonetically-detailed episodic traces in the lexicon with socially-indexed information, then we would expect that a specific race/ethnicity would prime a specific set of accent qualities. In contrast, listeners may be anticipating a nonnative accent when they encounter a minority race/ethnicity speaker, but not activating specific socio-phonetic categories. Under the latter account, listeners would essentially be "preparing for effortful listening" more broadly.

Lastly, we sought to examine not only how social information affects the overall perception of nonnative speech, but also how it affects perceptual adaptation to nonnative speech. Examining subjects' performance over time (i.e., across trials) raises the opportunity to examine immediate as well as sustained effects of social information. We expected improvement across the experiment for all conditions, and did not have any a-priori hypotheses about the rate of improvement differing across conditions. However, given our prediction that social information would have immediate, implicit effects on perception, we expected that the differences in performance may be observable as early as Trial 1.

2.1 Experiment 1

Pre-registration of Experiment 1, a replication of McGowan (2015), is available at https://osf.io/3dnwr. Data and analysis scripts for both Experiments 1 and 2 can be found at https://osf.io/bj8v9/files.

2.1.1 Method

Participants. Young adult participants were recruited from the Psychology Subject Pool at Washington University in St. Louis. Subjects were required to be native speakers of English with self-reported normal hearing and normal (or corrected-to-normal) vision, and to have no extensive exposure to Mandarin Chinese (i.e., cannot have studied Mandarin Chinese or lived with a native speaker of Mandarin Chinese). Subjects of any race were able to participate in the experiment, but our primary analyses focused on White subjects. This decision was informed by pilot data that indicated that a subject's race may influence their susceptibility to social priming effects (with larger effects seen in White subjects). Thus, we collected a racially representative sample of subjects, but report social priming effects for White subjects in our primary analyses and social priming effects for minority-race subjects as part of our exploratory analyses.

The target sample size was estimated with simulated dataset using the package simR (Green & MacLeod, 2016) in R (version 4.0.4). Effect sizes in the dataset were based on trends in McGowan (2015). We estimated greater than 80% power to find differences between groups with a sample of $N = 150$ (50 per group). Data collection ceased after 150 valid White subjects had completed the task. Altogether, 294 subjects participated in the experiment, and 57 were excluded from analyses for reasons as follows: non-normal hearing (seven), experience with Mandarin (40), failing attention-check trials (three), performing greater than three standard deviations below average (one), and self-reporting that their data should be excluded (e.g., reporting they were not paying attention during the task; six). This resulted in 150 White subjects (50 per group) and 87 subjects of other races/ethnicities (22 in the East Asian group, 28 in the White group, and 30 in the control group), as follows: Asian (24; includes South, Southeast, and East Asian responses), Black/African American (22), Hispanic/Latinx (five), Mixed Race/Ethnicity (26), and Native American (one). Hispanic/Latinx responses are summarized with the race data because multiple subjects provided this information within the open-response question about the subject's race. Two additional subjects opted to not provide their race/ethnicity and were excluded from both sets of data. All participants provided informed consent and were compensated with course credit as approved by the Washington University in St. Louis Institutional Review Board.

Materials. The same auditory stimuli were presented to all conditions. Targets were recordings of a Mandarin Chinese-accented female reading aloud 60 English Hearing-In-Noise-

Test (HINT; Nilsson, Soli, & Sullivan, 1994) sentences. These target sentences differed from the high and low predictability sentences used in McGowan (2015), and, coinciding with this decision, we did not investigate the effect of sentence predictability in the current experiment. Background noise samples were randomly extracted from a minute-long six-talker babble track. Babble files were created with 30 simple, meaningful sentences produced by three male and three female native speakers of American English (Bradlow and Alexander, 2007). Targets and babble were mixed at a signal-to-noise ratio of -2 dB, with target onset lagging 500 ms after the start of the babble. Two additional audio files were recorded by a different native speaker of American English for the attention-check trials. These files were recordings of the sentences "please type a single G" and "please type a single Q", and were presented without background noise.

For the visual stimuli (**Figure 2.1A**), pictures of a White and an East Asian female who were similarly rated for attractiveness, neutrality of expression, and high prototypicality of race were selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The Chicago Face Database's images are high quality photos cropped from the shoulders up. Photo subjects all wear the same gray t-shirt, directly face the camera, and maintain a neutral expression (i.e., mouth closed). The images used for the White and East Asian primes differed from those used in McGowan (2015). The visual stimulus for the control condition, a gray silhouette with no race information, was made to closely match the control image used by McGowan. During the experiment, one of the visual stimuli (depending on the subject's random assignment) was visible on the screen at all times.

Procedure. Participants completed the experiment online with Gorilla (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2018), and were strongly encouraged to use headphones for the task. After completing the consent process, subjects were given instructions for the task, which noted either that a photograph of the speaker would be shown during the task (for the White and East Asian prime conditions) or that no picture of the speaker was available (for the Control condition). Before beginning the task, subjects were presented with an example audio file to help them identify the target speaker among the babble speakers, and to allow them to adjust their volume as needed. The practice audio file was able to be played up to five times, and the correct response ("the boy fell from the window") was given to the subject.

During the task, the randomly-assigned picture was centered on the screen at all times. Each trial began with the target file playing over the headphones, and then a response box appeared on the screen for subjects to transcribe what they heard. Subjects were encouraged to take their best guess if they did not understand the speaker, and to use accurate spelling. After entering their response, there was a two-second delay before the next trial began. The task began with a three-trial practice section, and then transitioned to a section with 60 test trials (presented in a randomized order) and two attention-check trials. Attention-check trials were set to occur on trial 13 and 44 (approximately 33% and 66% of the way through the task). No breaks were administered during the task.

After the listening task, subjects completed a questionnaire regarding their language experience and demographic background. An additional question was included that allowed subjects to indicate if their data should be excluded for any reason (with examples such as "being too distracted"). Subjects were told that their response would not affect their course credit, and that answering honestly helped to ensure the validity of the study's data.

2.1.2 Results

Primary analyses of the White subject data are reported first, followed by exploratory analyses of priming differences based on subject race.

Generalized linear mixed-effects analyses. Data from the speech transcription task was modeled using the *glmer()* function from the *lme4* package in R (version 4.0.4). The predicted variable, transcription accuracy, was treated as a grouped binomial. That is, each trial of the task (for a given subject) corresponded to a single row of the dataframe in R, and the predicted variable of the model was two columns: correctly identified target words and missed target words. A logit link function was specified. Random effects included intercepts for subjects and items (i.e., target sentences). After visual inspection of the data indicated large improvement across the task, we decided to deviate from our pre-registered analysis plan. Rather than using a no-intercept model of the effect of prime (which would not have accounted for performance across trials), we decided to dummy-code the effect of prime (reference level: East Asian) and include an effect of trial and the interaction between prime and trial. A second set of models, in which the effect of prime was dummy-coded with Control as the reference level, was also constructed to directly compare the Control and White conditions. To assist with model convergence, the effect of trial was scaled with the *scale()* function in R.

The overall effect of prime did not improve model fit ($\chi^2(2) = 4.09$, $p = .13$). However, model estimates indicated a significant difference between the East Asian and White prime conditions ($\beta = -0.38$, $p = .04$), with poorer performance by subjects assigned to view the White prime (**Figure 2.1B**). The East Asian and Control primes and the White and Control primes did not differ significantly (both p's > .05).

The effect of trial and the interaction between trial and prime both significantly improved model fit ($\chi^2(1) = 572.44$, $p < .001$; $\chi^2(2) = 16.94$, $p < .001$, respectively). As shown in **Figure**

2.1C, performance improved across trials overall ($\beta = 0.64$, $p < .001$), but at a significantly faster rate for the Control than the East Asian ($\beta = 0.27$, $p < .001$) or the White ($\beta = 0.16$, $p < .001$) conditions. The White condition also improved at a marginally faster rate than the East Asian condition ($\beta = 0.11$, $p = .09$), though this difference was not significant.

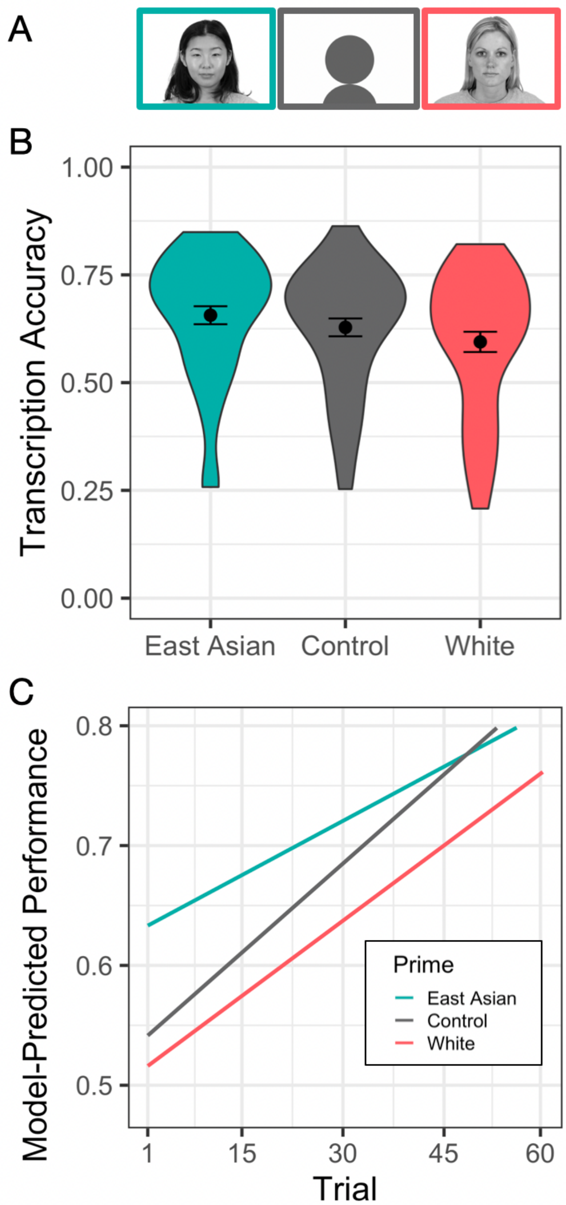

Figure 2.1. A) Pictures of the three priming images are shown. **B)** Effect of prime on transcription accuracy is visualized with violin plots and means with standard errors. The y-axis shows a summary of performance as proportion of words correctly transcribed. **C)** Performance across trials is shown with model fits. Y-axis shows the model-predicted probability of accurate perception.

We also examined the differences between the intercepts of each condition. For this line of inquiry, we created a version of the model containing only the effects of prime and trial, and set the numbering of trial to begin at zero. Thus, because the model-estimated differences between the levels of condition in this model represent differences at the intercept, they allowed us to compare performance at Trial 1. The trends at Trial 1 matched the conclusions from the overall effects: subjects assigned to the East Asian prime outperformed those assigned to the White prime (β = -0.39, p = .04), and no significant differences were detected when comparing either condition to the Control prime (both p's > .05).

Exploratory social priming in minority subjects. Specifications for the generalized linear mixed-effects models of the minority subject dataset matched those of the White subject dataset. Only the effect of trial improved model fit ($\chi^2(1)$ = 293.46, p < .001), indicating improvement across trials (β = 0.52). The main effect of prime did not improve model fit ($\chi^2(2)$ = 1.82, p = .40; **Figure 2.2**), and none of the model estimates indicated overall differences among the conditions (all p's > .05) or differences at the intercept (all p's > .05). The interaction between prime and trial also did not improve model fit ($\chi^2(2)$ = 1.34, p = .51).

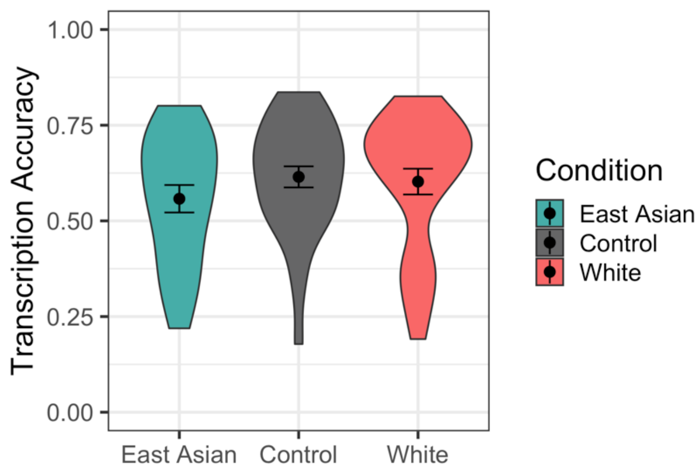

Figure 2.2. Violin plots show the distribution of minority subjects' performance. Points represent group means by priming condition with standard error bars.

We also compared the performance of the minority race/ethnicity subjects and the White subjects directly. Specifications for the generalized linear mixed-effects models matched those from earlier analyses. Our focus in this exploratory analysis was the effect of subject race (dummy-coded reference level: Minority) and the interaction between subject race and priming. When controlling for the effect of prime, overall performance by White and minority race/ethnicity subjects did not differ ($\beta = 0.20$, $p = .14$). The interaction between subject race/ethnicity and prime did not improve model fit overall ($\chi^2(2) = 4.33$, $p = .11$). However, model estimates indicated that the difference in performance for the East Asian and White conditions was flipped in direction depending on subject race/ethnicity ($\beta = -0.66$, $p = .04$). Minority race/ethnicity subjects performed worse when presented with the East Asian prime than the White prime, while White subjects performed better. The interaction when comparing the East Asian and Control prime groups and the Control and White prime groups was not significant (p's > .05).

2.1.3 Interim discussion

In Experiment 1, we replicated a finding from McGowan (2015), demonstrating that race information can prime White listeners during perception of Mandarin-accented English speech. Subjects who were randomly assigned to view an East Asian woman's face performed better during the task overall than subjects randomly assigned to view a White woman's face. Additionally, our novel analysis of performance across trials indicated that this priming effect is rapid – significantly affecting performance as early as Trial 1 of the experiment.

Subjects assigned to the control condition (i.e., to view a silhouette with no race information) did not perform significantly differently from the East Asian or the White priming conditions. However, examination of performance across trials provided a key insight into the performance of the control group subjects: Compared to the East Asian and White groups, the control group improved at the fastest rate. In other words, the control group showed the largest improvement across the task, beginning close to the level of performance of the White group and eventually reaching a level of performance that matched the East Asian group.

When accounting for change in performance across the experiment, it appears that the East Asian prime benefited listeners, and—importantly—that this benefit occurred immediately (at Trial 1). In other words, listeners appear to be accommodating the nonnative accent before the auditory signal is even encountered. These results support the hypothesis that linguistic representations encode meaningful social- and accent-related information, and link these representations to social (racial) categories. Viewing the East Asian face appears to activate socio-phonetic categories that facilitate more accurate perception of a Mandarin accent.

In contrast, the cost associated with viewing a White prime appears to be prolonged across the experiment. The difference in the rate of improvement (slope) for the control group versus the White group serves as the strongest evidence that the White prime is interfering with perceptual adaptation to the Mandarin accent. It may be the case that the White prime is activating socio-phonetic categories that are incongruent with the Mandarin accent, thus interfering with perceptual adaptation. However, another possibility (which is not mutually exclusive from the former) is that subjects are meta-cognitively aware of the incongruency, and this is affecting how they engage with and perform during the listening task. For example, subjects assigned to the White prime may (correctly) assess that they are being lied to about the image representing the speaker, and may consequently exert less effort during the listening task. Another explanation for the cost effect associated with the White prime is that subjects are surprised by the perceived 'incongruency' between the speaker's accent and race. However, this explanation would primarily explain an initial cost effect, not necessarily a prolonged cost effect.

We also conducted an exploratory analysis to determine whether subject race/ethnicity impacted the strength and/or direction of the social priming effect. Results indicated that the priming effect trended in the opposite direction for minority race/ethnicity subjects as compared to White subjects (i.e., poorer performance when viewing an East Asian prime than a White prime). However, our sample size for the minority subject group was smaller, and collapsed across multiple races/ethnicities. This analysis is best interpreted as an examination of the effect of "Whiteness" on the social priming effect. We anticipated that these priming effects may be stronger (or, perhaps, only present) in White subjects. Our exploratory analysis indicated that this may be the case, and motivates our focus on White subjects for all subsequent experiments. The aims of Experiments 2-4 are to determine whether these social priming effects occur for other nonnative accents (generalizability), the degree of their specificity, their relationship with implicit racial biases, and their impact on perceptual adaptation. We begin by addressing these questions in the population of subjects for which the replicated priming effect appears to be strongest. In future work, however, we believe that it will be crucial to revisit the effect of subject race and ethnicity on social priming.

2.2 Experiment 2

In Experiment 2, we added five random assignment conditions to the three examined in Experiment 1, which examined the effects of primes (Control, East Asian, and White) on the perception of Mandarin accent. First, we expanded the design to examine another nonnative accent: Arabic-accented English. Mirroring the design of Experiment 1, we included Control, *Middle Eastern*, and White prime conditions, for which we predicted similar trends as seen with Mandarin accent (i.e., demonstrating the *generalizability* of priming effects). Additionally, to examine the *specificity* of priming effects, for the Mandarin accent we added a Middle Eastern prime condition, and for the Arabic accent we added an East Asian prime condition. Altogether, this created a two (accent) by four (prime) design for Experiment 2.

2.2.1 Method

Participants. As in Experiment 1, subjects included young adults recruited from the Psychology Subject Pool at Washington University in St. Louis. All subjects were native speakers of English with normal hearing and normal (or corrected-to-normal) vision, and did not have extensive exposure to Mandarin Chinese or Arabic (screened based on randomly assigned condition). For Experiment 2, we specifically recruited White subjects. This decision was motivated by the exploratory analysis of race in Experiment 1, which indicated that these priming effects may vary by subject race and/or be more prominent in White subjects.

Given some of the marginal results of Experiment 1, we determined that power to detect the effect of priming may need to be increased in Experiment 2. To determine the sample size for Experiment 2, we used the data from Experiment 1 and the package simR (Green &

MacLeod, 2016). We estimated that by expanding the sample size to approximately 80 subjects per group, we would have greater than 80% power to find an effect of prime. We assumed that this sample size would be appropriate for the novel Arabic accent conditions as well. Data from Experiment 1 was included in Experiment 2 to conserve resources; thus, only 30 additional subjects were needed for the Mandarin accent conditions with the Control, East Asian, and White primes.

Altogether, 564 subjects participated in Experiment 2 (in addition to the 150 subjects from Experiment 1), and 74 were excluded from analyses and replaced for reasons as follows: reporting that they were not White (38; includes mixed-race White responses), non-normal hearing (six), experience with Mandarin (eight) or Arabic (five), failing attention-check trials (five), performing greater than three standard deviations below average (four), experiment error (one), and self-reporting that their data should be excluded (e.g., reporting they were not paying attention during the task; four). An additional three subjects were excluded because the number of subjects per condition accidentally exceeded 80 per group in some cases; timestamps of participation were used to exclude the last subjects to participate in the conditions with too many subjects. The final sample size was $N = 640$ (80 subjects per group). All participants provided informed consent and were compensated with course credit as approved by the Washington University in St. Louis Institutional Review Board.

Materials. The same 60 Hearing-In-Noise-Test (HINT; Nilsson, Soli, & Sullivan, 1994) sentences were used for both Experiments 1 and 2. A female Arabic-accented speaker of English was recorded reading aloud the target sentences. The speaker (who resided in Israel) was recorded remotely, resulting in slightly more background noise than in the Mandarin speaker's files. To reduce noise in the audio files, a noise-reduction procedure was applied with Adobe Audition (Build 13.0.13.46).

The same background noise samples used for the Mandarin speaker (sourced from a minute-long six-talker babble track) were mixed with the Arabic speaker's recordings. The Mandarin speaker was presented at a signal-to-noise ratio (SNR) of -2 dB in Experiment 1. For the Arabic speaker, it was clear from initial inspection of the files that -2 dB SNR would be too difficult. Thus, we conducted a brief pilot with 10 subjects in which the Arabic speaker was presented at 0 dB SNR. From this data, we determined that at 0 dB SNR the Arabic speaker was fairly well-matched to the performance levels in Experiment 1, and decided to use this SNR for the Arabic accent conditions in Experiment 2.

For the visual stimuli, we used the same images as in Experiment 1 for the Control, East Asian, and White prime (sourced from the Chicago Face Database; Ma et al., 2015). For the Middle Eastern prime, we were unable to select an image from the Chicago Face Database (CFD), which only contains photos of Asian, Black, Latinx, and White faces. Instead, the Middle Eastern prime used in Experiment 2 was a modeling headshot of a Middle Eastern woman wearing a hijab purchased from the website *iStock.com* (Getty Images). The image was selected because it was similar to the CFD images: the photo was cropped at the shoulders, the woman was directly facing the camera, and her facial expression was neutral (**Figure 2.3**). We conducted a pilot (12 subjects) which indicated that the image of the Middle Eastern woman was similarly rated for attractiveness, neutrality of expression, and prototypicality of race as the CFD images.

Figure 2.3. Priming images used for the East Asian, Control, White, and Middle Eastern random assignment conditions (in order from left to right).

Procedure. Procedures for Experiment 2 matched those of Experiment 1.

2.2.2 Results

Analyses of the Mandarin accent conditions are reported first, followed by analyses of the Arabic accent conditions, and finally comparisons across accent.

Analyses of Mandarin accent conditions. Data from the Mandarin accent conditions was modeled using the *glmer()* function with a logit link function, following the same process as Experiment 1. The predicted variable, transcription accuracy, was a grouped binomial: one column containing the number of correctly identified target words and another containing the number of missed target words. Random effects included intercepts for subjects and items (i.e., target sentences). The effect of prime was dummy-coded as: Control (reference level), East Asian, Middle Eastern, White. Trial was renumbered from 0 to 59 (i.e., differences at the intercept represent differences at Trial 1), and then scaled to assist with model convergence.

The overall effect of prime did not improve model fit ($\chi^2(3) = 2.79$, $p = .42$), and none of the conditions differed significantly from the control group (all *p*'s > .05; **Figure 2.4A**). The effect of trial did significantly improve model fit ($\chi^2(1) = 1142.60$, $p < .001$), as did the interaction between prime and trial ($\chi^2(3) = 34.03$, $p < .001$). Model estimates indicated that subjects assigned to the Control condition improved across the task at a faster rate than subjects assigned to any other condition (all p's < .001; **Figure 2.5A**). None of the intercepts of the priming conditions significantly differed from the Control condition (all p's > .05).

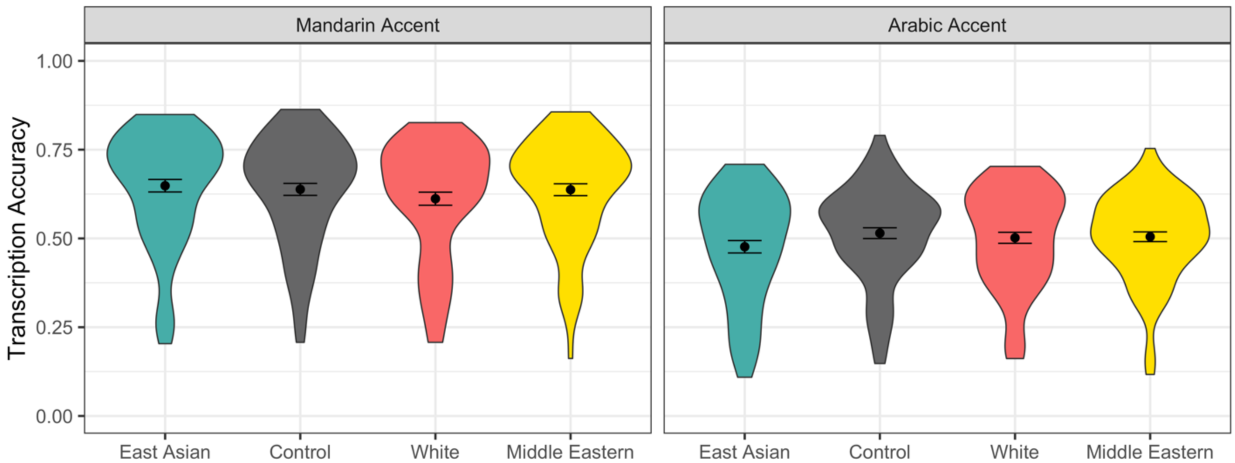

Figure 2.4. Violin plots with mean points and standard error bars show transcription accuracy (proportion of words correctly transcribed) for each priming condition within the Mandarin (left) and Arabic (right) accent conditions.

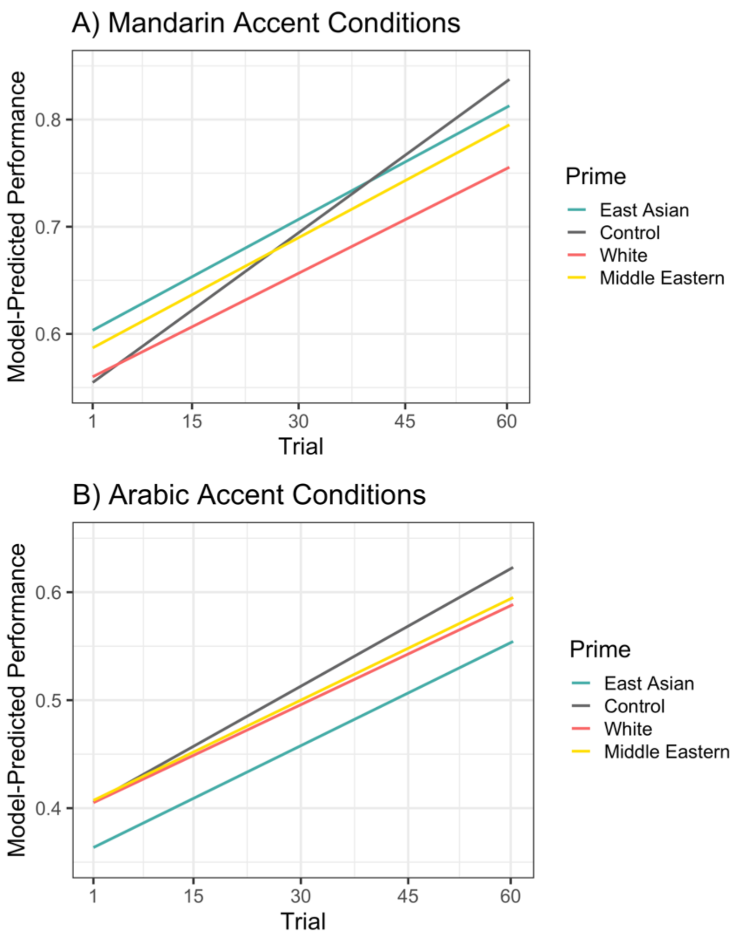

Figure 2.5. Model-predicted fit lines show probability of accurate sentence transcription at each trial across the task. Top graph (A) shows the Mandarin accent condition and bottom graph (B) shows the Arabic accent conditions. Y-axes span equal intervals but contain different ranges for each accent condition.

The model estimates were next reexamined with dummy-coding that made the White condition the reference level for the effect of prime. None of the other conditions significantly differed from the White condition (all *p*'s > .05), nor did their intercepts (all *p*'s > .05). The interaction between prime and trial did not reveal any differences in slopes between the White and East Asian or the White and Middle Eastern conditions. Lastly, the Middle Eastern condition was set as the reference level to determine whether subjects assigned to the East Asian prime differed from those assigned to the Middle Eastern prime. These conditions did not differ overall, at their intercepts, or in their slopes (all *p*'s > .05).

Analyses of Arabic accent conditions. The same analysis process applied to the Mandarin accent conditions was next applied to the Arabic accent conditions. Random effects included intercepts for subjects and items, and fixed effects included prime, trial, and the interaction between prime and trial. The effect of prime used the same initial dummy-coding scheme: Control (reference level), East Asian, Middle Eastern, White. Trial was once again renumbered from 0 to 59 (such that differences at the intercept represent differences at Trial 1), and then scaled to assist with model convergence.

Log-likelihood model comparisons indicated that the overall effect of prime did not improve model fit ($\chi^2(3) = 3.70$, $p = .30$). None of the priming groups differed significantly from the Control group (all *p*'s > .05; **Figure 2.4B**), although subjects assigned to the East Asian prime performed marginally worse than those assigned to the Control prime ($\beta = -0.22$, $p = .07$). The effect of trial significantly improve model fit ($\chi^2(1) = 744.11$, $p < .001$), and the model estimate indicated improvement across the task ($\beta = 0.46$, $p < .001$). The interaction between prime and trial, however, did not significantly improve model fit overall ($\chi^2(3) = 3.57$, $p = .31$). As shown in **Figure 2.5B**, subjects assigned to the Control prime did improve at a slightly faster rate than those assigned to the other primes, but this difference was only marginally significant for the Control versus the White prime group ($\beta = -0.08$, $p = .09$) and non-significant for the other two contrasts (other *p*'s > .10). None of the intercepts of the priming conditions significantly differed from the Control condition (all *p*'s > .05), although the intercept for the East Asian prime group was lower, and this difference was marginal ($\beta = -0.22$, $p = .07$).

Next, we altered the dummy-coding such that the White condition was the reference level for the effect of prime, and reexamined the model estimates. No conditions significantly differed from the White condition (all *p*'s > .05), nor did their intercepts (all *p*'s > .05). The interaction between prime and trial did not reveal any differences in slopes between the White and East Asian or the White and Middle Eastern conditions. Lastly, the Middle Eastern condition was set as the reference level to determine whether subjects assigned to the East Asian prime differed from those assigned to the Middle Eastern prime. These conditions did not differ overall, at their intercepts, or in their slopes (all *p*’s > .05).

Comparisons across accent conditions. The data from the Arabic and Mandarin accent conditions was combined into one dataset for a total of $N = 640$. Random effects included intercepts for subjects and items. Prime, trial, and accent, and the interactions between prime and accent and between trial and accent, were included as fixed effects. The effect of prime used the following (initial) dummy-coding scheme: Control (reference level), East Asian, Middle Eastern, White. The effect of accent was also dummy-coded, such that the Arabic accent was the reference level. Trial was renumbered from 0 to 59 (such that differences at the intercept represent differences at Trial 1), and then scaled to assist with model convergence.

The fixed effects of accent ($\chi^2(1) = 99.58$, $p < .001$) and trial ($\chi^2(1) = 1590.4$, $p < .001$) each significantly improved model fit. Although our intention was to match the overall intelligibility of the two accented speakers, model estimates indicated that subjects assigned to the Arabic-accented speaker had significantly lower transcription accuracy than subjects assigned to the Mandarin-accented speaker ($\beta = 0.62$, $p < .001$). As in the previous analyses, the effect of trial demonstrated that subjects improved across the task ($\beta = 0.46$, $p < .001$). The interaction between accent and trial also significantly improved model fit ($\chi^2(1) = 8.04$, $p = .005$), and indicated that perceptual adaptation to the accent was more rapid and greater in the Mandarin accent conditions than the Arabic accent conditions ($\beta = 0.07$, $p = .005$).

The overall effect of prime did not significantly improve model fit ($\chi^2(3) = 1.94$, $p = .59$). None of the priming conditions differed from the control condition (all *p*'s > .05). The interaction between prime and accent was also non-significant ($\chi^2(3) = 4.49$, $p = .21$), and model estimates did not reveal any differences between priming conditions based on accent (all *p*'s > .05).

Next, the White condition was set as the reference level in the dummy-coding for the effect of prime. No conditions significantly differed from the White condition overall (all *p*'s > .05). However, the interaction between prime and accent did reveal a difference between the White and East Asian priming conditions by accent; specifically, the model estimate indicated that subjects assigned to the Mandarin-accented speaker and the East Asian prime outperformed their White prime counterparts, while subjects assigned to the Arabic-accented speaker and the White prime outperformed their East Asian prime counterparts ($\beta = 0.34$, $p = .04$; see **Figure 2.4**). As a final step, the Middle Eastern condition was set as the reference level in the dummy-coding for prime. This contrast did not reveal any additional differences of the prime conditions overall or by accent (all *p*'s > .05).

2.3 Discussion

In Experiment 2, we expanded upon the replication of McGowan (2015) in order to investigate the generalizability and specificity of social priming effects. Unlike the results of Experiment 1, we did not find a priming effect for Mandarin-accented English speech; subjects shown an East Asian prime did not perform significantly better than subjects shown a White prime (although trends remained in the same direction). The inconsistent outcomes between Experiments 1 and 2 indicate that caution should be exercised when drawing conclusions about the effect of social primes on the perception of nonnative-accented speech. It may be the case that the size of this social priming effect is smaller than previously anticipated, or that there are additional factors that need to be taken into account.

A consistent finding between Experiments 1 and 2 was that subjects assigned to the view a silhouette with no race information (i.e., the control condition) when listening to Mandarin accent improved at the fastest rate. Subjects primed with an East Asian face appear to have an initial advantage over Control group subjects, but by the end of the experiment Control group subjects perform just as well or better than East Asian prime subjects. It may be the case that the East Asian prime activates socio-phonetic categories that facilitate more accurate perception of a Mandarin accent, allowing listeners to begin perceptually adapting to the nonnative accent before the auditory signal is even encountered. The silhouette prime does not provide this advantage, but it also does not inhibit adaptation to the accent.

One of the aims of Experiment 2 was to investigate the generalizability of social priming effects for nonnative accent. Modeled after the design of Experiment 1 (and McGowan, 2015), subjects were presented with Arabic-accented English speech and randomly assigned to one of the following three priming conditions: a Middle Eastern face, a silhouette image (control), or a White face. None of the three conditions differed overall or in their slopes. These results indicate that the social priming effect observed by McGowan (2015) and in Experiment 1 with Mandarin-accented English may not generalize to Arabic-accented English. However, one notable difference between the Mandarin and Arabic accent conditions in the current study was the overall difficulty of the transcription task. The Mandarin speaker was approximately 63% intelligible on average, while the Arabic speaker was approximately 50% intelligible on average. This difference in intelligibility occurred despite presenting the Arabic-accented sentences at an easier signal-to-noise ratio during the speech transcription task. The source of the difference in intelligibility may be attributable to the unique qualities of each speaker and each accent, or to lower quality of the Arabic accent recordings (which were created remotely due to the COVID-19 pandemic). Thus, while our data indicates that social priming did not occur for the Arabic accent in the current study, this line of investigation may be worth revisiting with a less challenging accented speaker or signal-to-noise ratio, or in a systematic investigation of the effects of task difficulty on social priming effectiveness.

In Experiment 2, we also examined the specificity of the social priming effect by introducing a fourth priming condition for each of the two accents. For the Mandarin accent this was a Middle Eastern priming condition, and for the Arabic accent this was an East Asian priming condition. Analyses of the Mandarin accent conditions did not reveal any significant differences between the Middle Eastern priming condition and the other priming conditions. Similar to the East Asian prime, trends in data suggested that subjects assigned to the Middle Eastern prime may have performed better than those assigned to the White prime. For the Arabic accent conditions, subjects assigned to the East Asian prime performed poorer than subjects in the other random assignment conditions, although this difference was only marginally significant when comparing against the Control group. Thus, the outcomes of the Mandarin-Middle Eastern and the Arabic-East Asian conditions appear to be at odds. We had posited two competing a-priori hypotheses for these conditions:

1) Social primes have a high degree of specificity, such that listeners associate specific races/ethnicities with specific accents. Thus, viewing a minority race/ethnicity prime that is uncommon for a given nonnative accent (e.g., a Middle Eastern face with a Mandarin accent) will incur a similar cost to listening performance as found for White primes.

2) Social primes are not highly specific; rather, listeners broadly associate minority races/ethnicities with nonnative accents. Thus, viewing a minority race/ethnicity prime that is uncommon for a given nonnative accent (e.g., a Middle Eastern face with a Mandarin accent) will not incur a cost to listening performance – and may even provide a benefit.

Trends in the Mandarin accent data appear to support Hypothesis 2, while trends in the Arabic accent data appear to support Hypothesis 1. Given the marginal and inconsistent results of the current study, this research question will need to be revisited in future research. However, our data do provide an important insight, and suggest that a third hypothesis ought to also be considered:

3) Social primes may vary in their specificity. Some accents may be associated with specific races/ethnicities, while others may not be. Additionally, individual listeners' unique experiences may shape the specificity and strength of these race/ethnicity-accent associations.

Despite the inconsistent outcomes between Experiments 1 and 2, the overall results of this study do indicate that social information (speaker race) can affect the perception of nonnative accent. Using a matched-guise design, we found that transcription accuracy for sentences presented in noise can be facilitated or hindered by the image of a face. However, this social priming effect may not occur consistently across nonnative accents. Indeed, a more nuanced model that accounts for individual listener experiences may be needed when examining the links between racial and sociophonetic categories. For example, the White listeners in our sample may have had more experience with East Asian nonnative accent varieties than Middle Eastern nonnative accent varieties. If this was the case, it would explain why the East Asian prime had a differential effect on the two accent types (i.e., a benefit for Mandarin accent and a cost for Arabic accent) while the Middle Eastern prime appeared to have a fairly benign effect on both (i.e., no strong benefit or cost for either accent). In future work, mapping the social networks and experiences of listeners may help to account for individual differences in the strength of these social priming effects.

3. Experiment 3: The Relationship Between Implicit Associations and the Social Priming of Speech

The primary aim of this study was to improve our understanding of the mechanism supporting social priming during speech perception. Based on our findings in Experiment 1, in which race information had an immediate effect on the perception of nonnative speech (i.e., at the first trial of the task), we concluded that social priming effects are both rapid and automatic. These results are in line with an exemplar model of speech perception, in which the lexicon contains episodic representations of words and phonemes. These episodic traces, or “exemplars”, are not only phonetically-detailed, but are also linked to social information (i.e., social characteristics of the speaker). Thus, social information, such as a speaker’s race, primes socio-phonetic categories that can either facilitate or inhibit accurate perception of a given accent. In the case of Experiment 1 (see also McGowan, 2015), viewing an East Asian woman’s face during perception of Mandarin-accented English in babble benefited listeners. However, other work examining perception of *native*-accented English in babble has found a cost associated with presentation of East Asian (specifically, Chinese) faces (Babel & Russell, 2015). These complementary findings indicate that racial and ethnic priming images affect perception of native versus nonnative accents types differently.

Another hypothesis derived from an exemplar model account of social priming is that socio-phonetic links should vary across listeners based on their unique experiences. If this is the case, we would expect that individuals with stronger associations between given racial and accent categories would have larger social priming effects. A wealth of research in social psychology has demonstrated that the strength of implicit associations between social categories (e.g., Old vs. Young) and constructs (e.g., Pleasant vs. Unpleasant) can be measured using Implicit Association Tests (IATs; Greenwald et al., 1998). Thus, to approach this hypothesis, in Experiment 3 we examine individual differences in implicit associations between racial/ethnic categories and categories such as *American* vs. *Foreign*. We discuss possible limitations of these selected IAT categories for capturing socio-phonetic associations in Chapter 5.

Notably, IATs have been used in prior work examining effects of race on speech perception. Yi and colleagues (2013) found that listeners derived less audiovisual benefit (i.e., an intelligibility benefit from adding the video signal of a talker to their corresponding audio signal) from Korean-accented English talkers than American-accented English talkers, and that this difference was related to implicit associations as measured with an IAT. However, in a direct replication with a larger sample size, McLaughlin and colleagues (2022) did not find evidence that IAT scores were related to reduced audiovisual benefit for Korean-accented English. In an experiment more similar to the current study, Babel and Russell (2015) attempted to examine individual differences in listeners' implicit associations using an IAT, and to determine whether these associations predicted social priming effects for native, Canadian-accented English. Their results did not indicate a relationship between IAT scores and social priming. In the current study, we aimed to improve upon this previous attempt by substantially increasing power to examine individual differences (details in Methods).

3.1 Experiment 3

In Experiment 3, we aimed to test whether there is a link between implicit associations and the effects of social primes on the perception of American-accented English. Additionally, we sought to examine social priming effects for two pairs of races/ethnicities: similar to prior work (Babel & Russell, 2015), we decided to examine priming for a White versus an East Asian prime, and, as a novel contribution, we decided to examine priming for a White versus a Latina prime. We refer to these two separate versions of the experiment as the East Asian-White and Latinx-White Cohorts, respectively. We posited the following hypotheses:

1. Race/ethnicity primes (images of a White, East Asian, vs. Latina woman) will affect the perception of American-accented English in babble. Specifically, subjects will have better transcription accuracy for the White prime than the East Asian or the Latina prime.
2. Individual listeners' implicit associations will be related to the size of their "social priming costs". That is, for a given listener, the size of the priming effect (e.g., performance following a White versus an East Asian prime) may be larger or smaller, and we expect that listeners with larger priming effects will have stronger implicit associations between these races/ethnicities and the constructs American vs. Foreign.
3. The relationship between social priming costs and implicit associations with the constructs Good vs. Bad will diverge from the relationship with American vs. Foreign associations. Specifically, we expect that Good vs. Bad implicit racial associations will not significantly predict social priming costs.
4. Explicit measures of listeners' attitudes and affect will also be examined. We predict that affect may be more negative following listening blocks with a minority race prime than a White prime. Additionally, we predict that ratings of status (e.g., intelligence) and solidarity (e.g., friendly) traits will be poorer following listening blocks with a minority race prime. Lastly, we predict that speakers' fluency will be rated poorer for listening blocks with a minority race prime. Altogether, we expect that explicit measures will

indicate bias in favor of the (supposedly) White speaker over the (supposedly) East Asian and Latina speakers.

Notably, one issue that we predicted may occur is an effect of counterbalancing on the speech transcription task and attitude measures. That is, after subjects enter the second block of the speech transcription task and become aware of the two speakers' (supposed) races/ethnicities, they may suspect that their racial biases are being examined. Further, because half of the subjects would be assigned to view the White prime first and the other half to view the East Asian (or Latina, depending on cohort assignment) prime first, we suspected that this may reduce our ability to detect an effect of social primes. Thus, in our pre-registered analysis plan we determined that if the effect of counterbalance was significant for the analyses of the social priming or attitude measures, then we would conduct an additional set of between-subject analyses with only data from the first block of the experiment (for which the same hypotheses would apply).

3.1.1 Methods

Pre-registration of Experiment 3 is available from https://osf.io/vdazs. Data and analysis scripts for the experiment can be found at https://osf.io/nd7wm/files. All procedures were approved by the Washington University Institutional Review Board.

Participants. Young adult subjects (age mean = 24.5; age range = 18-35) were recruited using the website Prolific to participate in Experiment 3 online. Inclusion criteria (set via Prolific's demographic filters) selected for White young adults between 18-35 years old, who reported English as their first and dominant language, currently residing in the United States and being of United States nationality, and having normal hearing and vision (or corrected-to-normal vision). We anticipated that some number of subjects would need to be excluded from the sample, and thus planned to recruit up to a maximum of 800 subjects (i.e., the maximum our budget allowed), or until 350 valid subjects participated in each cohort. In total, 778 subjects participated in the experiment, 74 of whom were excluded for one or more of the following reasons: failing to meet eligibility criteria (despite Prolific's pre-screening; six), failing the headphone screening (up to two attempts allowed; 33), self-reporting using speakers instead of headphones for any task (six), failing attention-check trials in the speech transcription task (two), performing greater than or equal to three standard deviations away from the group average in the speech transcription task (15), or self-reporting that their data should be excluded (12). The final *N* for each cohort was slightly above the target: 351 subjects in the East Asian-White Cohort and 353 subjects in the Latinx-White Cohort.

Sample size rationale. The target sample size was calculated with the aim of providing sufficient power to detect the relationship between IAT *D* scores and social priming costs. Using the function *pwr.r.test()* in R, we estimated power for an effect size of r = .20; this estimate was based on a meta-analysis of prior work in social psychology that has shown effect sizes of approximately $r = .20$ when examining correlations between IAT scores and intergroup behaviors (Greenwald, Poehlman, Uhlman, & Banaji, 2009). With 350 subjects per cohort, this analysis determined that there would be greater than 95% power to detect an effect size of $r = .20$ or greater. Additionally, in the case of a smaller relationship between implicit associations and social priming, this sample size would still provide 80% power to examine an effect as small as $r = .15$.

Materials. *Speech perception task.* Auditory stimuli were retrieved from SpeechBox (Bradlow, n.d.), and included recordings of two female American-accented native speakers of English reading aloud Hearing-In-Noise-Test sentences (Nilsson, Soli, & Sullivan, 1994;

Bradlow, Blasingame, & Lee, 2018). As in Experiments 1 and 2, samples of background noise were randomly extracted from a six-talker babble track created from 30 simple, meaningful sentences produced by three male and three female native speakers of American English (Bradlow & Alexander, 2007). The sentence targets and background babble were mixed at a signal-to-noise ratio (SNR) of -4 dB, with target onset lagging 500 ms after the start of the babble. Piloting of the two speakers indicated that at -4 dB SNR they were both approximately 60% intelligible.

For the attention-check trials, the same files from Experiments 1 and 2 were used. These files were recordings of the sentences "please type a single G" and "please type a single Q" read by an American English speaking female (i.e., a third female speaker), and were presented without background noise.

For the visual stimuli, the same images of a White and an East Asian female from Experiments 1 and 2 were selected for the East Asian-White cohort's primes. For the Latinx-White cohort's primes, an additional image of a Latina woman was selected from the Chicago Face Database (Ma et al., 2015). This prime was as closely rated as possible for attractiveness, neutrality of expression, and high prototypicality of race/ethnicity.

Implicit association tests (IATs). For both IATs and both cohorts, images of White, East Asian, and Latinx faces were selected from the Chicago Face Database using available metadata (Ma et al., 2015). Four male and four female faces were selected for each race/ethnicity, and were approximately matched (on average) for attractiveness, neutrality of expression, and high prototypicality of race/ethnicity. Additionally, we limited our selection of photos to individuals between 20-30 years of age to better match the ages of the speakers and pictured individuals in the speech perception task. The faces used in the IATs did not overlap with those selected for the speech perception task; this ensured that all faces in the IATs were equally novel. The same eight faces were used for the White condition in the East Asian-White cohort's and the Latinx-White cohort's IATs. All images were presented in color.

For the American-Foreign IAT, images of eight American scenes (e.g., the White House) and eight foreign scenes (e.g., the Taj Mahal) represented the constructs *American* and *Foreign*, respectively (photos from Yi et al., 2013). For the Good-Bad IAT, eight keywords were used for each construct in place of images (selected from Project Implicit's open materials; Xu et al., 2013). Good keywords included: happy, wonderful, love, pleasure, peace, joy, glorious, and laughter; Bad keywords included: hurt, agony, evil, nasty, terrible, horrible, failure, and awful.

Procedure. Subjects were first directed from Prolific to the experiment, which was hosted on Gorilla (Anwyl-Irvine et al., 2020). If subjects consented to participate, they then completed the following tasks (in order): a headphone screening, the speech transcription task (block one), the affect and attitudes questionnaire (block one), the speech transcription task (block two), the affect and attitudes questionnaire (block two), American-Foreign and Good-Bad IATs (order counterbalanced), and the demographic and language questionnaire. The procedures and tasks were the same for both cohorts, with the exception of which visual stimuli were presented. Subjects assigned to the East Asian-White cohort were presented with East Asian and White faces in the speech transcription task and IATs, and those in the Latinx-White cohort were presented with Latinx and White faces in these same tasks.

Headphone screening. The headphone screening was developed by Milne et al. (2020). The open-source version of the task (available in Gorilla) was used for Experiment 3. The screening leverages the perceptual phenomenon Huggins Pitch, which can only be detected when stimuli are presented dichotically (and, thus, should be unperceivable over loudspeakers). In the task, subjects are presented three noise bursts each trial, one of which contains a hidden tone within the noise. Subjects then make a forced-choice decision as to whether the pitch occurred during the first, second, or third noise burst. For the brief headphone screening, six trials are completed. In the present experiment, if subjects failed the screening, they were then given the opportunity to complete it a second time (after a reminder that headphones are required). Subjects who failed both attempts were immediately rejected/excluded from the study. However, approximately halfway through data collection a number of participant complaints indicated that the headphone screening was excluding users with Apple EarPods by mistake. At this point we changed the protocol to allow users who failed the screening to complete the task, and only excluded subjects who reported using computer speakers in the end-of-experiment questionnaire.

Speech transcription task. Before beginning each block of the speech transcription task, subjects were presented with an example audio file to help them identify the target speaker among the babble speakers. This file could be played up to ten times, and occurred without a visual prime. Subjects were told what the correct response was ("the boy fell from the window"), and to listen to the file until they were able to identify the target speaker. General instructions for the transcription task informed subjects that a photograph of the speaker would be shown during the task, that they should complete the task with their full attention and in a distraction-free environment, and that they should take their best guess when they didn't fully understand the speaker.

The racial prime for a given block was shown on the screen at all times during the transcription task. Each trial began with presentation of the target file in babble. After the file finished, a response box appeared on the screen for subjects to transcribe what they heard. A two second delay was inserted between trials. Each block contained 20 test trials and one attention-check trial presented in a randomized order. No breaks were administered during a given block. The presentation order of primes, and the combination of each prime and each of the two female speakers, was counterbalanced across subjects (four counterbalances total).

Affect and attitudes questionnaire. After each block of the speech transcription task, subjects completed a questionnaire assessing affect (emotional state) and attitudes toward the block's speaker (impressions of status, solidarity, and fluency). Questions for these assessments were based on Dragojevic & Giles (2016). Affect was assessed first via a series of six rating-scale questions in the frame: "How ___ are you feeling?" Negative valence prompts included irritated, annoyed, and frustrated, and positive valence prompts included interested, enthusiastic, and happy. Questions were presented in the same pseudorandom order for all subjects. The rating scale for the affect questions included five points, where "1 = not at all" to "5 = extremely". The current block's racial prime was not onscreen during the affect questions, because the questions were focused on the subject's emotional state.

For the attitudes questions, the racial prime of the block was shown at the top of the screen with the instruction: "Please indicate your impressions of the speaker shown above." All questions had the following frame: "How ___ is the speaker?" Status prompts included: intelligent, competent, smart, educated, and successful. Solidarity prompts included: nice, friendly, pleasant, honest, and sociable. Fluency prompts included: comprehensible, easy to understand, effortful to understand, and clear to understand. All questions were pseudo-randomly intermixed. The rating scale for the attitudes questions included seven points, where "1 = not at all" and "7 = extremely".

Implicit Association Tests (IATs). Procedures for the IATs followed the recommendations of Nosek, Greenwald, and Banaji (2005). The IAT is a response time sorting task containing seven blocks (**Table 3.1**). The key aspect of the IAT is how constructs are paired together in each block. In one block, subjects are told to sort images of White faces and American places into the same category, and images of East Asian faces and foreign places into a different category; then, in a different block these pairings are reversed (i.e., White faces with foreign places, East Asian faces with American places). Thus, by comparing response times across blocks, the IAT can measure associations between two sets of contrasted constructs.

Table 3.1

Block	**Number of trials**	**Left-key (d) response items**	**Right-key (k) response items**	**Function**
1	16	White	East Asian	Practice: Learn race dimension
2	16	American	Foreign	Practice: Learn attribute dimension
3	32	White + American	East Asian + Foreign	Race-attribute pairing 1 (Analyzed)
4	64	White + American	East Asian + Foreign	Race-attribute pairing 1 (Analyzed)
5	16	East Asian	White	Practice: Relearn race dimension
6	32	East Asian + American	White + Foreign	Race-attribute pairing 2 (Analyzed)
7	64	East Asian + American	White + Foreign	Race-attribute pairing 2 (Analyzed)

Note. Block order is counterbalanced across subjects. For half of the subjects, Blocks 1, 3, and 4 (pairing set 1) are swapped with Blocks 5, 6, and 7 (pairing set 2). The table above only shows pairings for the American-Foreign IAT, but procedures were identical for the Good-Bad IAT.

For the present IATs, the overall number of trials was increased 33.33%. Whereas usually each IAT would include approximately 180 trials (based on recommendations in Greenwald et al., 2003), the present study's IATs included 240 trials each. The number of trials in a given IAT can vary in part due to the number of target images per category (which can range from 2-10). However, in the present case, we made an intentional decision to increase the number of trials in

Blocks 4 and 7 in particular. Our aim was to increase the precision of the measure and, subsequently, improve our power to detect a relationship between IAT scores and performance in the speech perception task.

During each trial of the IAT, subjects are shown a single image or keyword and have to quickly sort into one of two categories. If subjects sort an item into the wrong category, a red 'X' appears on top of the item. The categories change each block (**Table 3.1**), and are always labeled in the left and right upper corners of the screen. To sort an item into the left category, subjects pressed the 'd' key, and to sort an item into the right category, subjects pressed the 'k' key. If subjects did not make a response within four seconds, the trial timed-out and the next stimulus was presented.

Block order was counterbalanced across subjects. For half of the subjects, Blocks 1, 3, and 4 (race-attribute pairing 1) were swapped with Blocks 5, 6, and 7 (race-attribute pairing 2). For example, for the American-Foreign IAT in the East Asian-White Cohort, half the subjects completed the White and American versus East Asian and Foreign pairings first, and the other half completed the East Asian and American versus White and Foreign pairings first.

Demographic and language background questionnaire. After a reminder that responses would not affect pay for the study, a series of questions asked subjects to report their age, gender, race, ethnicity, hearing status, nationality, and languages. A small set of additional questions was included for exploratory analyses: first, subjects were asked to indicate on a scale of 0% to 100%, what percentage of their social network included East Asian or Latinx individuals (depending on cohort assignment); second, subjects were asked to report in an open-response box what race/ethnicity they thought each of the speakers from the transcription task was (prime images were shown on the screen for these questions). Lastly, subjects were asked to report if they thought that there was any reason that their data should be excluded.

Data preparation. *Speech transcription task.* Transcription accuracy for each trial was calculated with the R package *autoscore* (Borrie, Barrett, & Yoho, 2019). Before using autoscore, keywords in the Hearing in Noise Test sentences were identified for each sentence. Specifically, determiners were excluded from scoring (e.g., “the”, “a”, “an”, “his”, “her”). The tallied number of correct versus incorrect (missed) keywords per sentence was used for analyses (details in Results).

Affect and attitudes questionnaire. A composite affect score was created for each subject and each block following the steps of Dragojevic and Giles (2016). First, the three positive (interested, happy, enthusiastic) and the three negative (annoyed, irritated, frustrated) emotions were each averaged to form positive and negative indices, respectively. Next, an affect balance score was calculated by subtracting the mean negative affect from the mean positive affect and adding a constant of 4 (to avoid negative values). Thus, the full (possible) range of the composite is from 0 (high negative affect) to 8 (high positive affect).

The attitudes measures were also summarized by subject and block. Rating scale selections (values between one and seven) were averaged to create separate composite measures of perceived status, solidarity, and fluency.

Implicit Association Tests (IATs). Only data from Blocks 3, 4, 6, and 7 was analyzed. Group-wide and individual IAT *D* scores were calculated for each IAT following Greenwald et al.’s (2003) guidelines. First, trials with latencies greater than 10,000 ms were excluded, and subjects with more than 10% of trials with latencies less than 300 ms were removed (this occurred for only one subject). Next, we calculated the mean latency for correct trials for each block, replacing error trial latencies with the sum of the original values plus an additional 600 ms. One pooled standard deviation was computed for all trials in Blocks 3 and 6, and another for all trials in Blocks 4 and 7. The average latencies for each block were used to calculate differences between Blocks 3 and 6 and between Blocks 4 and 7 (later blocks minus earlier blocks). These differences were divided by their respective pooled standard deviations, and then averaged to obtain *D* scores. Note that Blocks 3 and 6 are separated from Blocks 4 and 7 during this process because Blocks 3 and 6 are the initial trials of a given race and attribute pairing, and thus response times tend to be slower on average than during Blocks 4 and 7 (at which point subjects have typically gotten the hang of the new sorting rule).

3.1.2 Results

We first report separate analyses of the East Asian-White Cohort and the Latinx-White Cohort, and then exploratory analyses of the combined datasets.

East Asian-White Cohort. *Speech transcription task.* We used the *glmer()* function from the *lme4* package in R to fit the data from the speech transcription task. The predicted variable, transcription accuracy, was treated as a grouped binomial. That is, each trial of the task (for a given subject) corresponded to a single row of the dataframe in R, and the predicted variable of the models was two columns: correctly identified target words and missed target words. A logit link function was specified for the models. Random effects included intercepts for subjects and items, and random slopes for prime by subject.

Fixed effects in the model included: prime (dummy-coded levels: East Asian, White), American-Foreign *D* score, Good-Bad *D* score, affect, perceived status, perceived solidarity, perceived fluency, and counterbalance (dummy-coded levels: Counterbalance 1, Counterbalance 2), as well as interactions between prime and each additional predictor. Counterbalance 1 contained subjects who were presented with the East Asian prime block first, and Counterbalance 2 contained subjects who were presented with the White prime block first. Log-likelihood model comparisons were used to determine whether each fixed effect significantly improved model fit (summarized in **Table 3.2**). Most notably, the effect of prime did not significantly improve model fit ($\chi^2(1) = 0.01$, $p = .94$), contrary to our predictions (**Figure 3.1**).

Table 3.2

EFFECT	$\chi^2(1)$	*p*
Prime	0.01	.94
American-Foreign *D* score	2.05	.15
Good-Bad *D* score	0.57	.45
Affect	7.84	.005 **
Perceived status	3.49	.06
Perceived solidarity	6.86	.009 **
Perceived fluency	27.13	< .001 ***
Counterbalance	5.18	.02 *
Prime x American-Foreign *D* score	0.10	.75
Prime x Good-Bad *D* score	0.40	.53
Prime x Affect	0.43	.51
Prime x Status	0.10	.75
Prime x Solidarity	0.05	.82
Prime x Fluency	2.56	.11
Prime x Counterbalance	72.37	< .001 ***

Note. Summary of log-likelihood model comparisons for the East Asian-White Cohort analysis.
* = $p < .05$
** = $p < .01$
*** = $p < .001$

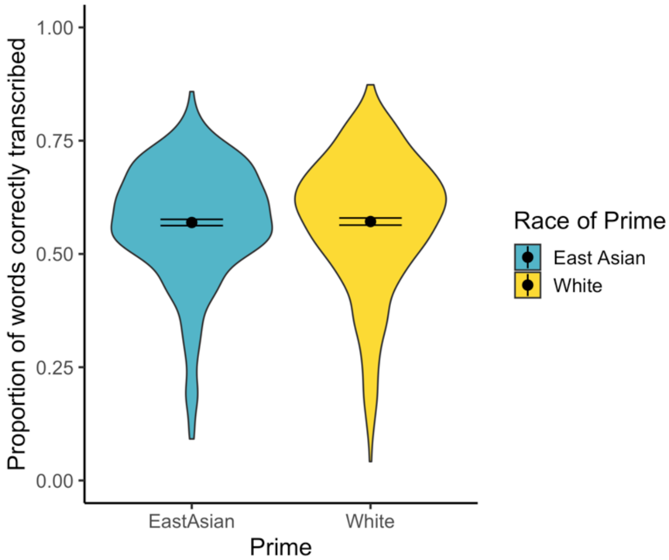

Figure 3.1. Violin plots, means, and standard error bars show the non-significant effect of prime (i.e., viewing an East Asian versus White face) on transcription accuracy. The y-axis is a summary measure of performance, showing the proportion of keywords transcribed accurately per sentence (averaged across trials).

Neither the individual differences in American-Foreign *D* scores nor in Good-Bad *D* scores improved model fit (both *p*'s > .05). The effect of perceived status also did not significantly improve model fit, though it was marginal ($\chi^2(1) = 3.49$, $p = .06$), but the effects of affect ($\chi^2(1) = 7.84$, $p = .005$), perceived solidarity ($\chi^2(1) = 6.86$, $p = .009$), and perceived fluency ($\chi^2(1) = 27.13$, $p < .001$) all significantly improved model fit. Model estimates indicated that subjects who performed better on the speech transcription task subsequently reported more positive affect ($\beta = 0.05$), and rated the given speaker higher for perceived solidarity ($\beta = 0.11$) and perceived fluency ($\beta = 0.15$).

The effect of counterbalance was also significant ($\chi^2(1) = 5.18$, $p = .02$), and indicated that subjects presented with the White prime block first performed better overall ($\beta = 0.16$; **Figure 3.2**). The only model interaction that significantly improved model fit was between prime and counterbalance ($\chi^2(1) = 72.37$, $p < .001$; all other *p*'s > .05). As visualized in **Figure 3.2**, this interaction is driven by improvement from Block 1 to Block 2: better performance on the White prime for Counterbalance 1 and better performance on the East Asian prime for Counterbalance 2.

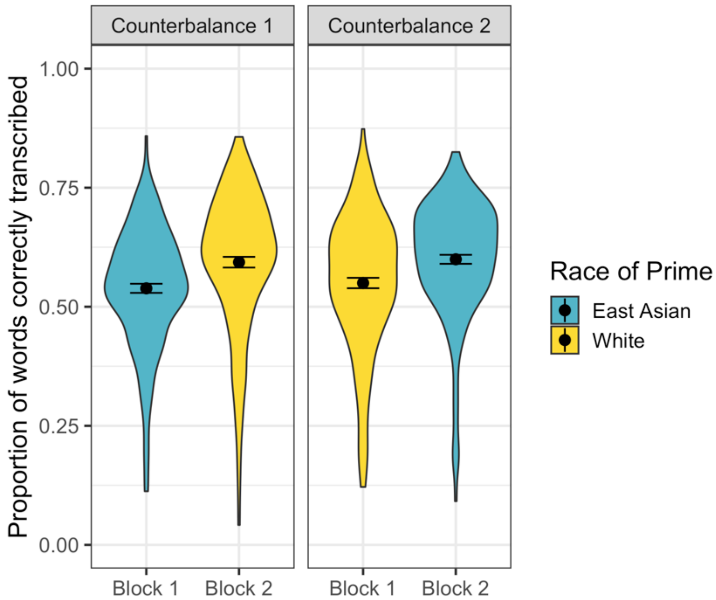

Figure 3.2. Violin plots with means and standard errors show the significant interaction between counterbalance and prime. The y-axis summarizes performance on the speech transcription task as proportion of words correctly perceived. Subjects in Counterbalance 1 were presented with the East Asian prime block first, while subjects in Counterbalance 2 were presented with White prime block first. The main effect of counterbalance was also significant, indicating better performance overall by subjects in Counterbalance 2.

We anticipated that the order of presentation of the race primes (counterbalancing) may affect social priming, and pre-registered an additional set of analyses for this scenario. Following the same analysis plan, we constructed a set of models comparing data only from the first block of the speech transcription task. This resulted in a between-subject analysis, with 174 subjects in the East Asian prime group and 177 subjects in the White prime group. This change required the random effect structure to be altered to include only random intercepts of subjects and items. Fixed effects for the analysis included: prime (dummy-coded levels: East Asian, White), American-Foreign *D* score, Good-Bad *D* score, affect, perceived status, perceived solidarity, perceived fluency, and interactions between prime and each of the other effects. Log-likelihood model comparisons are summarized in **Table 3.3**.

Table 3.3

EFFECT	$\chi^2(1)$	p
Prime	0.01	.91
American-Foreign *D* score	3.66	.06
Good-Bad *D* score	1.47	.23
Affect	7.83	.005 **
Perceived status	16.35	< .001 ***
Perceived solidarity	16.20	< .001 ***
Perceived fluency	18.27	< .001 ***
Prime x American-Foreign *D* score	0.17	.68
Prime x Good-Bad *D* score	3.52	.06
Prime x Affect	1.53	.22
Prime x Status	0.11	.75
Prime x Solidarity	1.26	.26
Prime x Fluency	0.24	.63

Note. Summary of log-likelihood model comparisons for the analysis of Block 1 data from the East Asian-White Cohort.
* = $p < .05$
** = $p < .01$
*** = $p < .001$

As was the case for the within-subject analysis, the effect of prime did not significantly improve model fit ($\chi^2(1) = 0.01$, $p = .91$). Thus, it does not appear that the effect of block was reducing our ability to detect an effect of social priming. However, one notable change that emerged in the between-subject comparison was the effect of American-Foreign *D* scores; when including individual differences in American-Foreign implicit associations, the fit of the model was marginally improved ($\chi^2(1) = 3.66$, $p = .06$). As shown in **Figure 3.3**, the model estimate was negative ($\beta = -0.21$), indicating that subjects with larger *D* scores performed poorer on Block 1 of the speech transcription task.

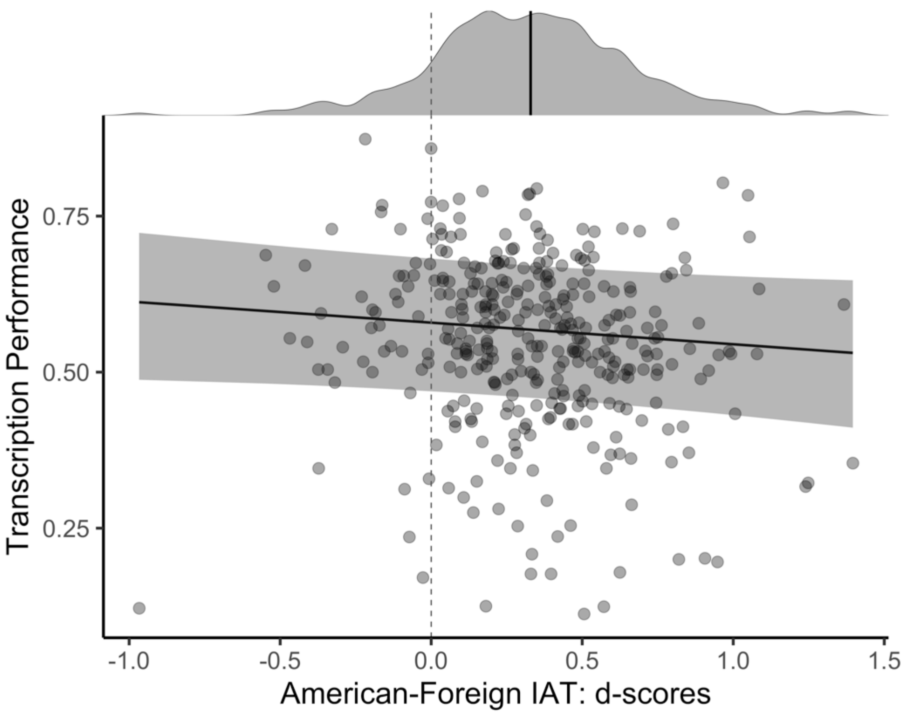

Figure 3.3. Visualization of the relationship between performance on Block 1 of the speech perception task (summarized as proportion of keywords correctly transcribed) and individual differences in American-Foreign IAT *D* scores for the East Asian-White Cohort. In lower plot, subjects are represented as individual points with a model-predicted fit line overlaid with standard error, and, in upper plot, a density distribution of the IAT *D* scores is shown with the group mean as a solid line. Values on the x-axis greater than zero (see dashed line) indicate stronger associations between White and American and between East Asian and Foreign, while values lower than zero indicate stronger associations between East Asian and American and between White and Foreign.

The effect of perceived status also emerged as significant when examining the Block 1 data ($\chi^2(1) = 16.35$, $p < .001$). Unlike the other attitudes measures, which had positive model estimates, the estimate for perceived status was negative ($\beta = -0.25$). Thus, subjects who performed more poorly on the speech perception task subsequently rated the speaker higher for status traits (e.g., intelligence), but lower for solidarity (e.g., friendliness) and fluency (e.g., comprehensibility).

Outcomes for the interactions between prime and the other effects all remained the same, with the exception of the interaction between prime and Good-Bad *D* scores, which emerged as marginally significant ($\chi^2(1) = 3.52$, $p = .06$). However, upon further inspection of the data it became clear that this interaction was driven by outliers in the distribution of *D* scores (defined as greater than three standard deviations away from the mean). Removal of these values (five subjects total) changed the interaction between prime and Good-Bad *D* scores such that it was not marginal ($p = .70$). Thus, we did not examine this interaction further.

Variance inflation factors. As a last step, we checked the variance inflation factors (i.e., multi-collinearity) of the main effects in the within-subject model. In particular, we were concerned that, because *D* scores from the American-Foreign and Good-Bad IATs were significantly correlated ($r = 0.27$, $t = 5.37$, $p < .001$), the two factors may not be capturing unique sources of variance. **Table 3.4** summarizes the variance inflation factors for both the East Asian-White and Latinx-White Cohorts. For the East Asian-White Cohort, variance inflation factors for the American-Foreign and Good-Bad *D* scores were very low, and did not indicate an issue of multi-collinearity or need to examine the two effects in separate models. Only the effects of perceived status and perceived solidarity indicated a small-to-moderate correlation among predictors. However, we determined these variance inflation factors to be within an acceptable range.

Table 3.4

East Asian-White Cohort		**Latinx-White Cohort**	
Fixed effect	**VIF**	**Fixed effect**	**VIF**
Prime	1.02	Prime	1.01
AF *D* score	1.07	AF *D* score	1.12
GB *D* score	1.06	GB *D* score	1.12
Affect	1.16	Affect	1.23
Perceived status	2.79	Perceived status	2.90
Perceived solidarity	2.60	Perceived solidarity	2.46
Perceived fluency	1.42	Perceived fluency	1.58
Block	1.01	Block	1.03

Note. Variance inflation factors are shown for the East Asian-White and Latinx-White Cohorts. All values were determined to be within an acceptable range.

Affect and attitudes questionnaire. For each of the composite measures of affect and attitudes, we constructed linear mixed-effects models and used log-likelihood model comparisons to examine the effects of prime (dummy-coded levels: East Asian, White), counterbalance (dummy-coded levels: Counterbalance 1, Counterbalance 2), and the interactions between prime and counterbalance. Random intercepts were included by subject. Where interactions between the effect of prime and counterbalance were significant, follow-up linear models containing only data from Block 1 were constructed (as pre-registered).

For the model of affect: the effect of prime was non-significant ($\chi^2(1) = 1.08$, $p = .30$), the effect of counterbalance was non-significant ($\chi^2(1) = 0.12$, $p = .72$), but the interaction between prime and counterbalance was significant ($\chi^2(1) = 8.03$, $p = .005$). In the follow-up model containing only Block 1 data, no difference between the two priming conditions was found ($\beta = -0.12$, $p = .49$).

In the model of perceived status, prime significantly improved fit ($\chi^2(1) = 37.47$, $p < .001$), and the model estimate indicated that the White prime was rated lower for status than the East Asian prime ($\beta = -0.24$). The effect of counterbalance was not significant ($\chi^2(1) = 0.41$, $p = .52$), nor was the interaction between prime and counterbalance ($\chi^2(1) = 1.19$, $p = .27$).

The White prime was also rated significantly lower for solidarity than the East Asian prime ($\beta = -0.14$; $\chi^2(1) = 12.43$, $p < .001$). The effect of counterbalance did not significantly improve fit ($\chi^2(1) < .01$, $p = .99$), but the interaction between prime and counterbalance did improve fit significantly ($\chi^2(1) = 14.33$, $p < .001$). As shown in **Figure 3.4**, subjects assigned to Counterbalance 1 (which was shown the East Asian prime during Block 1) rated the East Asian prime higher for solidarity as compared to the White prime, while those assigned to Counterbalance 2 did not rate the two primes differently. However, the follow-up model containing only Block 1 data did not reveal a significant difference between the two priming conditions (β = -0.14, p = .16).

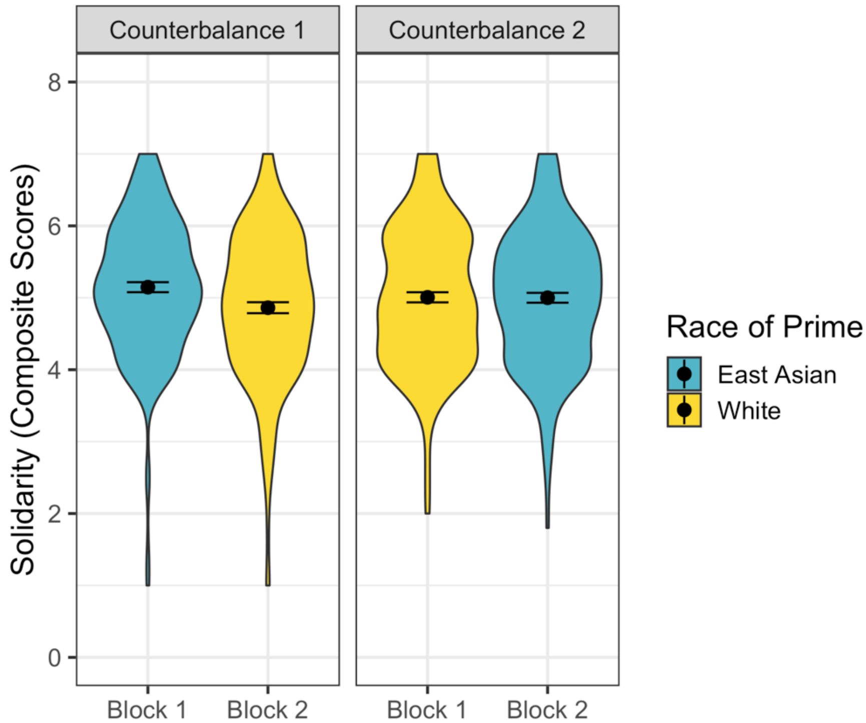

Figure 3.4. Violin plots show the distribution of solidarity ratings by block and counterbalance. Points represent means with standard error bars.

Lastly, perceived fluency was more negative following blocks with the White prime (β = -0.15; $\chi^2(1) = 9.36$, p = .002). Neither the effect of counterbalance ($\chi^2(1) = 0.59$, p = .44) nor the interaction of prime and counterbalance ($\chi^2(1) = 0.91$, p = .34) improved model fit.

Implicit Association Tests (IATs). Linear mixed-effects regression was used to examine the response time data from the American-Foreign and Good-Bad IATs. For both models, random effects included random intercepts by subject and random slopes by condition. Fixed effects included the effect of condition (ordered levels: congruent, incongruent), counterbalance, and the interaction between condition and counterbalance. In the IATs, blocks which combined White faces with American scenes (or Good keywords) and East Asian faces with Foreign scenes (or Bad keywords) were labeled as congruent, referring in this case to the expected biases of the subjects. Blocks which combined White faces with Foreign scenes (or Bad keywords) and East Asian faces with American scenes (or Good keywords) were labeled as incongruent. We predicted that, group-wide, response times would be slower for incongruent as compared to congruent conditions.

For the American-Foreign IAT, condition improved model fit ($\chi^2(1) = 146.06$, $p < .001$), and the model estimate indicated slower response times in incongruent blocks ($\beta = 88.53$). Counterbalance did not improve model fit ($\chi^2(1) = .003$, $p = .95$), but the interaction between condition and counterbalance was significant ($\chi^2(1) = 7.83$, $p = .005$). Subjects who completed the congruent block before the incongruent block had larger effects of condition than subjects who completed the incongruent block first ($\beta = 36.86$). Results of the Good-Bad IAT mirrored those of the American-Foreign IAT: condition improved fit ($\chi^2(1) = 53.72$, $p < .001$; $\beta = 69.22$), counterbalance did not improve fit ($\chi^2(1) = 2.39$, $p = .12$), and the interaction improved fit ($\chi^2(1) = 16.92$, $p < .001$; $\beta = 74.07$). **Table 3.5** summarizes the group-wide statistics of the IATs from both cohorts.

Table 3.5

East Asian-White Cohort				Latinx-White Cohort			
American-Foreign IAT		Good-Bad IAT		American-Foreign IAT		Good-Bad IAT	
d	α	*d*	α	*d*	α	*d*	α
0.34	0.71	0.20	0.74	0.33	0.64	0.24	0.64

Note. Summary statistics for the IATs are shown for both cohorts. Cronbach's alpha values were calculated within condition and then averaged.

Latinx-White Cohort. *Speech transcription task.* Model specifications matched those in the analysis of the East Asian-White Cohort. Fixed effects in the *glmer()* model included: prime (dummy-coded levels: Latinx, White), American-Foreign *D* score, Good-Bad *D* score, affect, perceived status, perceived solidarity, perceived fluency, and counterbalance (dummy-coded levels: Counterbalance 1, Counterbalance 2), as well as interactions between prime and each additional predictor. Counterbalance 1 contained subjects who were presented with the Latina prime block first, and Counterbalance 2 contained subjects who were presented with the White prime block first. Log-likelihood model comparisons were used to determine whether each fixed effect significantly improved model fit (summarized in **Table 3.6**). Similar to our findings in the analysis of the East Asian-White Cohort, the effect of prime did not significantly improve model fit ($\chi^2(1) < 0.01$, $p = .99$; **Figure 3.5**).

Table 3.6

EFFECT	$\chi^2(1)$	*p*
Prime	< 0.01	.99
American-Foreign *D* score	2.97	.08
Good-Bad *D* score	5.56	.02 *
Affect	9.46	< .001 ***
Perceived status	0.07	.79
Perceived solidarity	1.41	.23
Perceived fluency	42.87	< .001 ***
Counterbalance	3.68	.06
Prime x American-Foreign *D* score	1.01	.32
Prime x Good-Bad *D* score	3.05	.08
Prime x Affect	1.44	.23
Prime x Status	0.53	.47
Prime x Solidarity	1.01	.31
Prime x Fluency	0.05	.83
Prime x Counterbalance	74.92	< .001 ***

Note. Summary of log-likelihood model comparisons for the Latinx-White Cohort analysis.
* = $p < .05$
** = $p < .01$
*** = $p < .001$

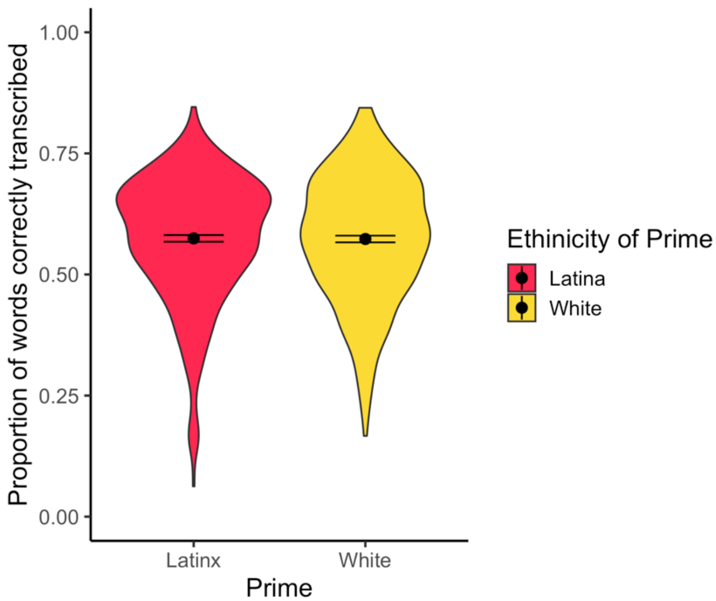

Figure 3.5. Violin plots, means, and standard error bars show the non-significant effect of prime (i.e., viewing a Latina versus White face) on transcription accuracy. The y-axis shows the proportion of keywords transcribed accurately per sentence (averaged across trials).

For the IATs, individual differences in Good-Bad *D* scores significantly improved model fit ($\chi^2(1) = 5.56$, $p = .02$). The model estimate ($\beta = -0.18$) indicated that subjects with larger *D* scores on the Good-Bad IAT performed poorer overall on the speech transcription task. Individual differences in American-Foreign *D* scores, however, only made a marginal (non-significant) contribution to model fit ($\chi^2(1) = 2.07$, $p = .08$). **Figure 3.6** shows the relationship between overall task performance and Good-Bad *D* scores. In this case, larger *D* scores indicate stronger associations between White faces and the construct Good and between Latinx faces and the construct Bad (smaller *D* scores indicate the opposite).

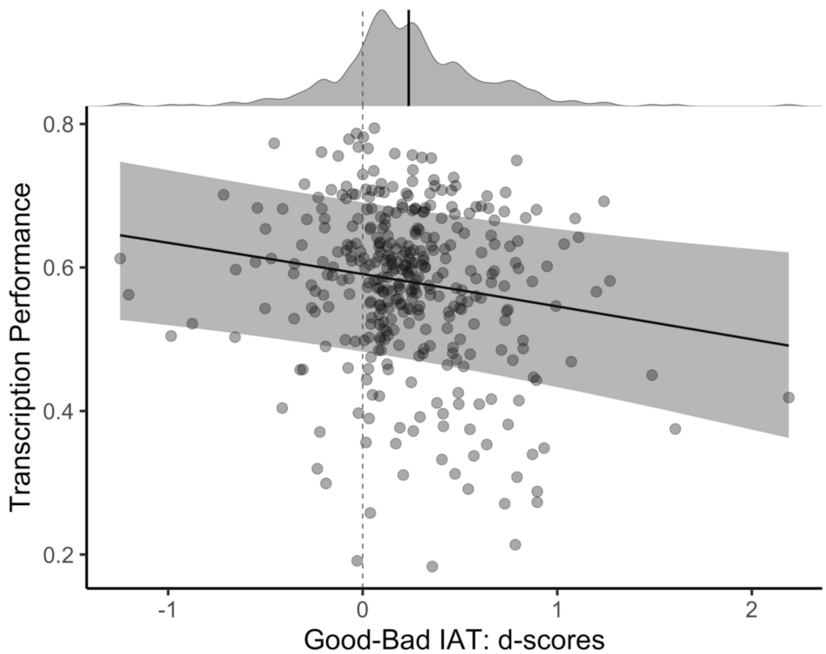

Figure 3.6. Visualization of the significant relationship between performance on the speech perception task (summarized as proportion of keywords correctly transcribed) and individual differences in Good-Bad IAT *D* scores for the Latinx-White Cohort. In lower plot, subjects are represented as individual points with a model-predicted fit line overlaid with standard error, and, in upper plot, *D* scores are summarized with a density distribution and solid line showing the group mean. Values on the x-axis greater than zero (see dashed line) indicate stronger associations between White and Good and between Latinx and Bad, while values lower than zero indicate stronger associations between Latinx and Good and between White and Bad.

Of the four affect and attitudes composites, only the effects of affect ($\chi^2(1) = 9.46$, $p < .001$) and perceived fluency ($\chi^2(1) = 45.87$, $p < .001$) significantly improved model fit. The effects of perceived status ($\chi^2(1) = 0.07$, $p = .79$) and perceived solidarity ($\chi^2(1) = 1.41$, $p = .23$) did not improve model fit. Model estimates indicated that subjects who performed better on the speech transcription task subsequently reported more positive affect ($\beta = 0.05$), and rated the given speaker higher for perceived fluency ($\beta = 0.18$). Lastly, the effect of counterbalance made a marginal improvement to model fit ($\chi^2(1) = 3.68$, $p = .06$), and the interaction between prime and counterbalance was significant ($\chi^2(1) = 74.92$, $p < .001$). As was the case for the East Asian-White Cohort, the model estimates (main effect $\beta = 0.12$, interaction $\beta = -0.54$) indicated a large improvement from Block 1 to Block 2 (**Figure 3.7**), and (marginally) better performance overall for subjects in Counterbalance 2, who were assigned to view the White face during Block 1.

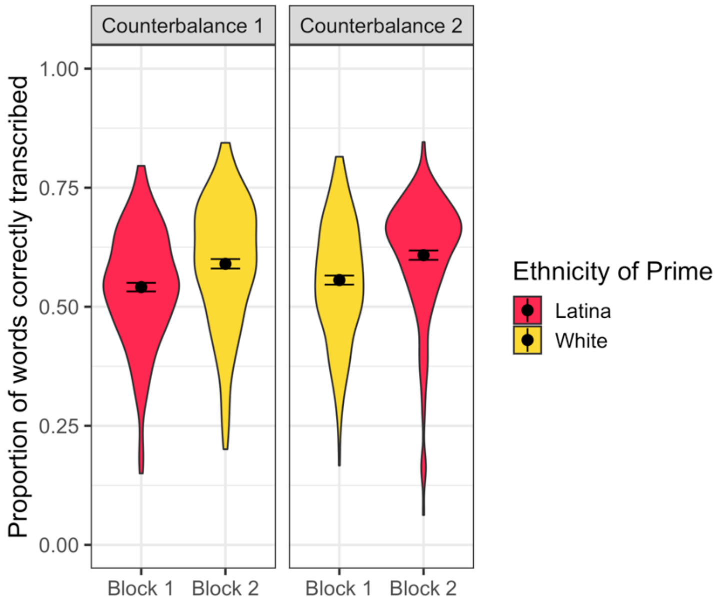

Figure 3.7. Violin plots with means and standard errors show the significant interaction between counterbalance and prime. The y-axis summarizes performance on the speech transcription task as proportion of words correctly transcribed. Subjects in Counterbalance 1 were presented with the Latina prime block first, while subjects in Counterbalance 2 were presented with White prime block first.

As was the case for the analysis of the East Asian-White Cohort, none of the other model interactions significantly contributed to fit (all *p*'s > .05). However, because of the interaction between prime and counterbalance, we created a model with data from the first block of the speech transcription task only. For this between-subject analysis, our sample included 178 subjects in the Latina prime group and 175 subjects in the White prime group. Random and fixed effects matched those in the between-subject analysis of the East Asian-White Cohort. Log-likelihood model comparisons are summarized in **Table 3.7**.

Table 3.7

EFFECT	$\chi^2(1)$	*p*
Prime	0.02	.88
American-Foreign *D* score	2.78	.10
Good-Bad *D* score	7.39	.007 **
Affect	3.10	.08
Perceived status	0.18	.67
Perceived solidarity	0.98	.32
Perceived fluency	17.77	< .001 ***
Prime x American-Foreign *D* score	1.83	.18
Prime x Good-Bad *D* score	0.13	.72
Prime x Affect	1.17	.28
Prime x Status	0.36	.55
Prime x Solidarity	0.84	.36
Prime x Fluency	0.16	.69

Note. Summary of log-likelihood model comparisons for the analysis of Block 1 data from the Latinx-White Cohort.
* = $p < .05$
** = $p < .01$
*** = $p < .001$

The results remained largely the same as for the within-subject analysis, with the exception of the effect of affect, which only marginally improved model fit ($\chi^2(1) = 3.10$, $p = .08$). The effect of prime did not improve model fit ($\chi^2(1) = 0.02$, $p = .88$). Individual differences in American-Foreign *D* scores also did not significantly (or marginally) improve fit ($\chi^2(1) = 2.78$, $p = .10$), while individual differences in Good-Bad *D* scores did ($\chi^2(1) = 7.39$, $p = .007$). Perceived status ($\chi^2(1) = 0.18$, $p = .67$) and solidarity ($\chi^2(1) = 0.84$, $p = .36$) did not improved fit, but perceived fluency did ($\chi^2(1) = 17.77$, $p < .001$). None of the interactions between prime and the additional measures were significant (all *p*'s > .05).

Variance inflation factors. As reported in **Table 3.4**, we checked the variance inflation factors (i.e., multi-collinearity) of the main effects in the within-subject model for the Latinx-White Cohort. This was motivated the correlated *D* scores from the American-Foreign and Good-Bad IATs ($r = 0.31$, $t = 6.18$, $p < .001$). As was the case for the East Asian-White Cohort, for the Latinx-White Cohort variance inflation factors for the American-Foreign and Good-Bad

D scores were very low, and did not indicate an issue of multi-collinearity or need to examine the two effects in separate models. Only the effects of perceived status and perceived solidarity indicated a small-to-moderate correlation among predictors, which we determined to be within an acceptable range.

Affect and attitudes. As in the East Asian-White Cohort analyses, we examined each of the composite measures of affect and attitudes with linear mixed-effects models and log-likelihood model comparisons. The fixed effects included prime (dummy-coded levels: Latina, White), counterbalance (dummy-coded levels: Counterbalance 1, Counterbalance 2), and the interactions between prime and counterbalance. By-subject random intercepts were included.

In the model of affect, prime did not improve model fit ($\chi^2(1) = 0.10$, $p = .75$), but the effects of counterbalance ($\chi^2(1) = 4.78$, $p = 0.03$) and the interaction between prime and counterbalance ($\chi^2(1) = 17.51$, $p < .001$) did improve fit. Model estimates indicated that subjects in Counterbalance 2, which was assigned to view the White prime during Block 1, reported more positive overall affective state ($\beta = 0.37$). Additionally, the significant interaction indicated that the effect of priming on affective state was reversed for each counterbalance (**Figure 3.8**). As for the speech perception task, a follow-up linear model containing data only from Block 1 (i.e., between-subjects) was constructed to determine the effect of priming on affective state. The model estimate confirmed that subjects assigned to view the White prime during Block 1 subsequently reported higher affect than those assigned to the Latina prime during Block 1 ($\beta = 0.35$, $p < .05$). Thus, before subjects are aware of the two priming conditions, the randomly assigned racial prime for Block 1 influenced their emotional state.

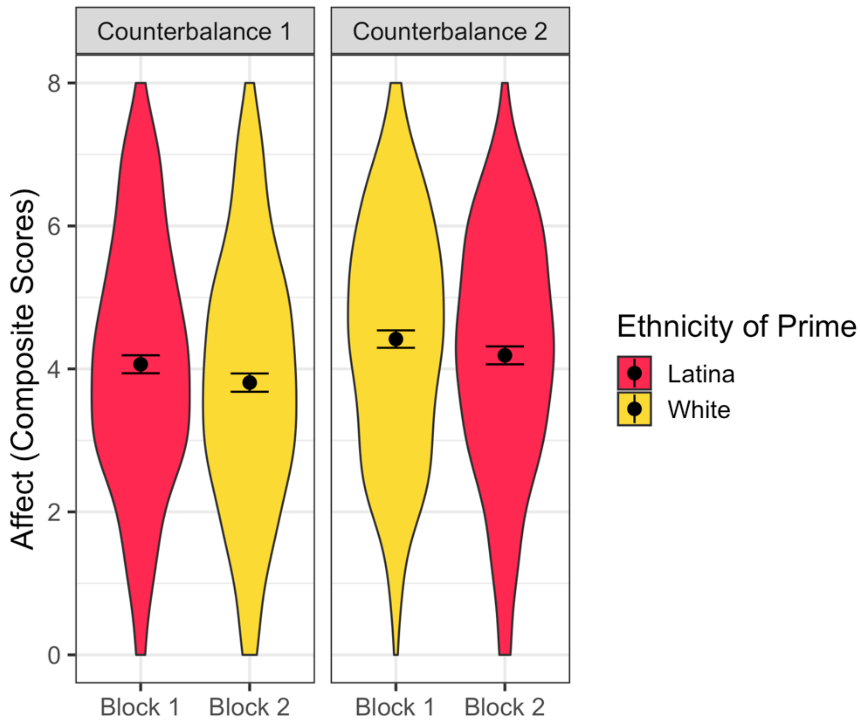

Figure 3.8. Violin plots show the distribution of affect ratings by block and counterbalance. Points represent means with standard error bars.

For the model of perceived status, log-likelihood model comparisons indicated: a significant effect of prime ($\chi^2(1) = 4.13$, $p = .04$), a non-significant effect of counterbalance ($\chi^2(1) = 0.73$, $p = .39$), and a non-significant interaction between prime and counterbalance ($\chi^2(1) = 1.65$, $p = .20$). Subjects rated the White prime lower on the measures of status than the Latina prime ($\beta = -0.08$).

There was no difference between solidarity ratings for the White and Latina prime ($\chi^2(1) = 0.01$, $p = .91$). The effect of counterbalance was not significant ($\chi^2(1) = 0.48$, $p = .49$), but the interaction between prime and counterbalance was significant ($\chi^2(1) = 22.20$, $p < .001$). As shown in **Figure 3.9**, this interaction was driven by an effect of block; as compared to Block 1, subjects rated the speaker lower for solidarity following Block 2. The follow-up model containing only Block 1 data did not reveal a difference between the group assigned to the White versus the Latina prime ($\beta = 0.08$, $p = .47$).

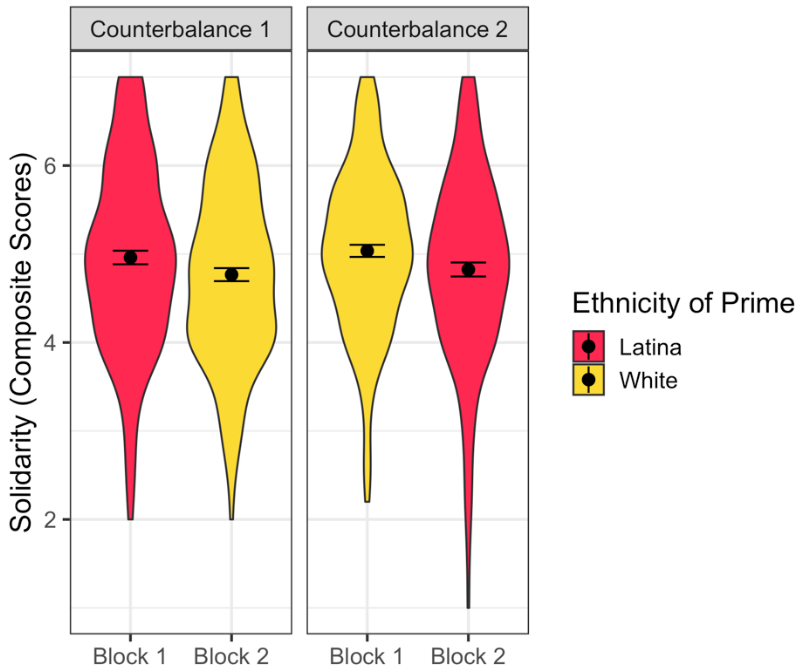

Figure 3.9. Violin plots show the distribution of solidarity ratings by block and counterbalance. Points represent means with standard error bars.

Lastly, there was a marginal effect of prime for the model of perceived fluency ($\chi^2(1) = 3.70$, $p = .05$). The direction of the estimate once again indicated more negative ratings for the White prime than the Latina prime ($\beta = -0.10$). The effect of counterbalance ($\chi^2(1) = 0.49$, $p = .48$) and the interaction between prime and counterbalance ($\chi^2(1) = 0.17$, $p = .68$) did not significantly improve model fit.

Implicit Association Tests (IATs). Random and fixed effect structures of the linear mixed-effects models of the IAT response time data for the Latinx-White Cohort matched those for the East Asian-White Cohort. For the American-Foreign IAT, the effect of condition was significant ($\chi^2(1) = 126.25$, $p < .001$), and the model estimate indicated slower response times in incongruent blocks ($\beta = 89.56$). The effect of counterbalance was not significant ($\chi^2(1) = .41$, $p =$

.52), but the interaction between condition and counterbalance was significant ($\chi^2(1) = 4.62$, p = .03). Log-likelihood model comparisons for the Good-Bad IAT mirrored those of the American-Foreign IAT: condition improved fit ($\chi^2(1) = 70.22$, $p < .001$; $\beta = 73.90$), counterbalance did not improve fit ($\chi^2(1) = 1.36$, $p = .24$), and the interaction improved fit ($\chi^2(1) = 17.45$, $p < .001$; β = 69.44). For both IATs, the significant interactions indicated that subjects who completed the congruent block before the incongruent block had larger effects of condition than subjects who completed the incongruent block first.

Exploratory analyses. *Combined analysis.* Based on the analyses of the Block 1, it appeared that the insignificant effects of prime were, nonetheless, consistent in direction and size across cohorts. Thus, we decided to examine the effect of prime within Block 1 in a combined dataset, with the levels of East Asian and Latina combined into a single level: Minority. Our aim was to increase power to detect a difference between primes ($N = 705$; 352 White, 353 Minority). Random effects included intercepts by subject and by item. Only the fixed effect of prime was entered into the model. The effect of prime did not improve model fit ($\chi^2(1) = 0.02$, $p = .89$), reaffirming the null outcomes of the primary analyses.

Reliability of speech transcription task. Cronbach's alpha of the speech transcription task was calculated for the East Asian-White and Latinx-White cohorts, respectively. Separate alpha values were calculated by counterbalance and block to reduce effects of prime and block on the reliability measure (eight values total), and then averaged. The speech transcription task showed acceptable reliability for both cohorts ($\alpha = .84$ and $\alpha = .78$, respectively).

Open response race/ethnicity categorization. Open response answers for the race/ethnicity categorization of the prime images were manually coded as belonging to one of the following categories: American Indian or Alaskan Native, Asian (included all South, Southeast, and East Asian responses), Black or African American, Latinx or Hispanic, Middle Eastern, Native Hawaiian or Other Pacific Islander, White or Caucasian, or Response Not Sortable (RNS; e.g., responses such as “American”). For the East Asian-White Cohort, 96.2% of responses for the White prime image were coded as White or Caucasian and the remaining 3.8% of responses were coded as RNS. For the East Asian prime image, 99.2% of responses were coded as Asian and the remaining 0.8% of responses were coded as RNS. For the Latinx-White Cohort, the White prime was coded as White or Caucasian for 98.9% of responses, Latinx or Hispanic for 0.3% of responses, and RNS for 0.8% of responses. The Latina prime was coded as Latinx or Hispanic for 94.6% of responses, Asian for 2.4% of responses, Middle Eastern for 0.8% of responses, Native Hawaiian or Other Pacific Islander for 0.3% of responses, White or Caucasian for 1.1% of responses, and RNS for 0.8% of responses.

3.1.3 Discussion

In Experiment 3, we found little evidence of social priming on the perception of American-accented English speech. Transcription accuracy for sentence-length materials presented in noise was not significantly different for East Asian or Latina primes than it was for a White prime. There was only one indication of an effect of race/ethnicity on performance during the speech perception task: an effect of counterbalancing. This effect indicated that subjects assigned to the view the White prime during the first block performed better on the task overall than those assigned to view the East Asian or Latina primes first. This effect of counterbalancing was significant in the East Asian-White Cohort, and marginal (p = .06) in the Latina-White Cohort. While the effect of counterbalancing indicates a cost associated with the minority race/ethnicity primes, direct examination of the effect of prime does not. Additionally, even when examining the effect of prime using only the Block 1 data (i.e., between-subjects) it is non-significant, indicating that the effect of prime is not necessarily being obscured by improvement from Block 1 to Block 2. Altogether, the results of Experiment 3 provide little evidence to indicate that speaker race/ethnicity information affects the intelligibility of American-accented English speech for White American-English listeners.

These findings conflict with those of Babel & Russell (2015), in which presentation of Chinese-Canadian speakers' faces resulted in poorer transcription accuracy than control (fixation cross) trials. However, the design of our study differed from the design of Babel & Russell (2015) in multiple ways: 1) We used a matched-guise paradigm in place of pictures of the actual speakers; 2) We blocked our task, presenting only two different speaker (and prime) blocks; and, 3) We collected data online, sampling a population of American listeners across the United States (whereas Babel & Russell examined a local population of listeners in Vancouver, BC, Canada). We address each of these topics in turn.

The matched-guise paradigm used in Experiment 3 was adapted from the design used by McGowan (2015) and in Experiments 1 and 2. This paradigm has two primary benefits: the combination of images and audio can be completely counterbalanced, and images can be carefully selected to match for qualities such as attractiveness, prototypicality of race/ethnicity, and facial expression. Both of these aspects allow for greater experimental control. However, one potential limitation of our implementation of the matched-guise paradigm in Experiment 3 was that a single prime was used to represent each race/ethnicity. Coinciding with this decision, the task was blocked such that subjects completed 20 trials with one prime and talker combination, and then 20 trials with another prime and talker combination (counterbalanced across subjects). In contrast, Babel & Russell (2015) used stimuli from 12 native Canadian-accented English talkers (6 Chinese-Canadian, 6 White-Canadian), the primes were pictures of the actual speakers, and the stimuli and priming condition were presented in a randomized order.

Additionally, it is possible that by sampling a population of subjects across the United States, we inadvertently introduced noise into our sample. Some communities may have larger nonnative speaker populations, and others smaller, resulting in different socio-phonetic mappings in the native listeners. Indeed, Babel & Russell (2015) discuss that in the Metro Vancouver area (from which they sampled their listeners) stereotypes of Asian Canadians as second language speakers of English are prominent, which may account for why they found a priming effect even with a much smaller sample of subjects than in the current study. Additionally, it is possible that subjects from the Metro Vancouver area have greater overall exposure to Asian faces.

The secondary aim of Experiment 3 was to examine the potential relationship between implicit racial associations and social priming of speech perception. Subjects completed two implicit association tests, examining associations with the constructs American vs. Foreign and the constructs Good vs. Bad. In the East Asian-White Cohort data, a marginal relationship between performance on Block 1 of the speech transcription task and *D* scores from the American-Foreign IAT indicated that listeners with stronger White + American and East Asian + Foreign associations had poorer transcription accuracy. However, the interaction between the IAT *D* scores and priming was non-significant. A similar marginal relationship for the American-Foreign IAT emerged in the Latinx-White Cohort data. Additionally, for the Good-Bad IAT, a significant relationship emerged: listeners with stronger White + Good and East Asian + Bad associations had poorer transcription accuracy for the task overall (and when examining only Block 1). Once again, however, individual differences in IAT scores did not interact with the effect of the social primes for the Latinx-White Cohort.

Thus, it appears that subjects with stronger implicit racial associations (in the directions noted above) performed more poorly on the speech perception task, but that this was unrelated to the race of the prime presented on the screen. While it is possible that subjects with stronger biases are also worse at speech-in-noise tasks, we suggest another possibility: The IATs and speech task may be related via a third variable. Relationships between individual differences in executive cognitive control (e.g., inhibition, working memory, and task switching) and IAT scores have been documented (Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010; Ito et al., 2015), as have relationships between individual differences in executive control (particularly working memory; Zekveld, Rudner, Johnsrude, & Rönnberg, 2013; Yeend, Beach, & Sharma, 2019) and the ability to understand speech in noise (evidence for a role of inhibition is mixed; Janse, 2012; McLaughlin et al., 2018). In the case of IATs, individuals with greater executive function skills tend to have smaller IAT *D* scores, because they are better able to control their responses during the IAT. Thus, the correlation between the IAT and the speech transcription task may actually represent a relationship between subjects' cognitive abilities and listening performance. Subjects with greater executive control would be expected to perform better at the speech transcription task and also have smaller IAT *D* scores, resulting in the negative relationship found.

We also assessed subjects' affect and attitudes toward each speaker after each block of the task. For the affect and attitudes measures, we used rating scale questions from Dragojevic and Giles (2016; e.g., "How enthusiastic are you feeling?" and "How intelligent is the speaker?") and then created composites of affect (negative vs. positive valence), perceived status, perceived solidarity, and perceived fluency. The prime picture remained on the screen during the attitudes portion of the questionnaire. For the East Asian-White Cohort data, priming did not appear to influence subjects' affect, but it did impact ratings of perceived status, solidarity, and fluency. In contrast to our predictions, all three of the attitude measures favored the East Asian prime over the White prime. Given that the Good-Bad IAT indicated (group-wide) implicit biases in favor of White over East Asian faces, it is possible that subjects gave more favorable explicit ratings of the East Asian prime in an effort to appear unbiased. However, it should also be noted that because only a single prime was used to represent each race, differences between the two women's faces may have also affected subjects' ratings; the images were selected to match as closely as possible for attractiveness and neutrality of expression, but there are other features that are unrelated to racial differences that may have deviated in the two images.

For the Latinx-White Cohort data, affect was impacted by counterbalancing such that subjects assigned to view the White prime in Block 1 had more positive affect than subjects assigned to view the Latina prime (both overall and when comparing Block 1 responses between-subjects). Similar to the outcomes of the speech transcription task, affect was not directly affected by the priming manipulation. For the attitudes measures, a similar trend emerged as seen in the East Asian-White Cohort data: the Latina prime was rated more favorably for perceived status and perceived fluency. The difference in perceived solidarity was non-significant. As suggested above, it may be the case that subjects intentionally gave better ratings for the Latina prime on these explicit measures in order to avoid appearing biased.

Relationships between the affect and attitudes measures and performance on the speech perception task were also estimated. For both cohorts, subjects who performed better on the speech perception task subsequently reported more positive affect, and gave higher ratings for fluency. Status ratings were not related to speech perception task performance for either cohort.

In the East Asian Cohort, subjects who performed better on the speech perception task also gave higher ratings for perceived solidarity.

Altogether, the results of Experiment 3 provide some evidence that race information may impact perception of American-accented English speech, but this evidence is inconsistent. In some cases, listeners with stronger implicit racial biases performed worse during the speech perception tasks, but this was unrelated to the race of the speaker presented on the screen during the task. Additionally, the social prime presented in the first block of the task had an impact on performance in the speech task overall, but the priming effect itself was non-significant. These inconsistencies indicate that social priming effects for native-accented speakers, if they exist, may be substantially smaller than previously estimated.

4. Experiment 4: The Effect of Social Primes on Multi-talker Perceptual Adaptation

While it may be difficult to process a nonnative accent initially, listeners can rapidly adapt to (or "accommodate") unfamiliar nonnative-accented speech (Clarke & Garrett, 2004; Brown, McLaughlin, Strand, & Van Engen, 2020). Notably, perceptual adaptation occurs for a single talker as well as multiple talkers, and perceptual learning can generalize to a novel talker of the same accent (Bradlow & Bent, 2008) or a different accent (Baese-Berk et al., 2013). Altogether, work examining accommodation of nonnative accent indicates that listeners not only adapt to a given nonnative speaker, but that they can learn the systematic deviations present across multiple speakers with nonnative accent and generalize that learning.

In Experiment 4, our aims were to examine the effect of social primes on perceptual adaptation to multiple nonnative-accented talkers, and on generalizability of perceptual learning to a novel nonnative talker. To this aim, we created a multi-talker training session with three Cantonese-accented English talkers, and a post-test session with a fourth (novel) Cantonese-accented English talker. The training and post-test sessions were both transcription-based tasks, in which subjects listened to sentences read aloud by the talkers and then typed what they heard into a response box. Priming images were only presented in the training session, and not at test. Feedback was not provided. We predicted the following:

1. Subjects assigned to view East Asian primes during the Cantonese-accented English, multi-talker training session will have better transcription accuracy during training than subjects assigned to view White primes. Additionally, this benefit of the East Asian primes will be immediate (significant at Trial 1) as well as sustained. These outcomes

would indicate that the East Asian primes provide an overall benefit for perception of nonnative, Cantonese accent.

2. For the test session, subjects who were trained with the East Asian primes will outperform subjects who were trained with the White primes. The East Asian prime group will also do better at test than subjects in the Control group (who receive no training session). We do not have a prediction as to whether the White prime group will do better than the Control group. This outcome would indicate that perception training for nonnative accent is benefitted by 'congruent' race information, and improves generalizable perceptual learning.

4.1 Experiment 4

Pre-registration of Experiment 4 is available at https://osf.io/dr4m5 and data and analysis scripts can be found at https://osf.io/dr4m5/files/.

4.1.1 Method

Participants. Young adult subjects (age mean = 26.6; age range = 18-35) were recruited using the website Prolific to participate in Experiment 3 online. Separate postings were used to recruit the control group subjects and the training group subjects, because the control group's session was substantially shorter (approximately 10 minutes, as compared to 25 minutes for the training group subjects). Both groups were paid at the same rate approved by the Washington University Institutional Review Board, rounding up to the nearest 15-minute increment: $2.50 for the control group, $5.00 for the training groups. Inclusion criteria (set via Prolific's demographic filters) selected for White young adults between 18-35 years old, who reported English as their first and dominant language, currently residing in the United States and being of United States nationality, and having normal hearing and vision (or corrected-to-normal vision). Our recruitment plan and budget allowed up to a maximum of 300 subjects (100 per group), although we did recruit three replacement subjects to even out the final number of subjects per group after exclusions (90 per group). In total, 303 subjects participated in the experiment, 33 of which were excluded for one of the following reasons: failing to meet eligibility criteria (despite Prolific's pre-screening; 14), self-reporting using speakers instead of headphones for any task (two), failing attention-check trials in the speech transcription task (four), performing greater than or equal to three standard deviations away from the group average in the speech transcription task (10), self-reporting that their data should be excluded (one), or software error (two).

Materials. During the training session, both training groups were presented with the same auditory stimuli featuring three female Cantonese-accented speakers of English. In the post-test session, a fourth female Cantonese-accented speaker was presented for all random assignment groups. All targets were Hearing-In-Noise-Test sentences (HINT; Nilsson et al., 1994), retrieved from the SpeechBox corpus (Bradlow, n.d.). The four target speakers were selected after piloting eight Cantonese-accented female speakers to determine their relative intelligibility levels; the four talkers selected were similarly intelligible when presented in -2 dB signal-to-noise ratio (SNR) speech-shaped background noise. The speech-shaped noise used in piloting and in the final experiment was created using the target speaker files with Praat (version 6.1.10; Boersma & Weenink, 2021). Background noise extended before and after the target sentence for 500 ms. As in Experiments 1-4, three additional audio files were also recorded by a native speaker of American English for the attention-check trials. These files were recordings of the sentences "please type a single G", "please type a single Q", and "please type a single X", and were presented without background noise. The first two catch trials were presented during the training session, and the third was presented during the post-test session.

For the visual stimuli, pictures of three White and three East Asian females who were similarly rated for attractiveness, neutrality of expression, and high prototypicality of race were selected from the Chicago Face Database (Ma et al., 2015). The Chicago Face Database's images are high quality photos cropped from the shoulders up. Photo subjects all wear the same gray t-shirt, directly face the camera, and maintain a neutral expression (i.e., mouth closed). Subjects randomly assigned to the East Asian priming condition were only shown the three East Asian female faces, while those assigned to the White priming condition were only shown the three White female faces. No prime was presented during the test session.

Procedure. Subjects were redirected from Prolific to the experiment hosted on Gorilla (Anwyl-Irvine et al., 2020). Subjects in the training groups completed the following tasks (in order): a headphone screening, the training session, the post-test session, and the demographic and language questionnaire. The training session was the same for both random assignment training groups, with the exception of the race of the visual stimuli presented. Subjects recruited for the control group completed the following tasks (in order): a headphone screening, the post-test session, and the demographic and language questionnaire.

Headphone screening. The headphone screening was designed by Milne et al. (2020). Subjects are played three bursts of noise and have to determine which contained a hidden tone. Notably, in our prior work (Experiments 3 and 4) this screening has had issues with mistakenly excluding subjects who use Apple EarPods. For this reason, we did not use the screening to determine which subjects to exclude. However, many Prolific subjects drop out of the experiment at the headphone screening (likely because they are not wearing headphones and believe they cannot deceive it), which still makes it a useful screening tool. A self-report question about headphones at the end of the task was used to exclude subjects from analyses instead.

Training session. Before beginning the training session, subjects were instructed to do the speech transcription task in a distraction-free environment, listen closely to each sentence, take their best guess if they did not understand the speaker, and use correct spellings (to the best of their abilities). Subjects were also reminded of the headphone requirement at this time. Additional instructions for the training session informed subjects that a photograph of each given speaker would appear on the screen before each trial and that they should not close their eyes during the task.

During the training session, each prime co-occurred with the same talker's voice consistently (20 trials per prime/talker), appearing on screen 2000 ms before the audio stimulus began. After the audio stimulus finished, a response box appeared on the screen for subjects to transcribe what they heard. A two second delay was inserted between trials. The session contained 62 trials total, including the two attention-check trials which were set to occur approximately one-third and two-thirds of the way through the task. The presentation order of the target sentences was randomized across subjects. Halfway through the experiment, subjects were told that they could take a short break before continuing the task.

Post-test session. Instructions for the post-test session matched those for the training session, with the exception of the instructions about the images of the speakers. Subjects were told that during the post-test session there would be nothing on the screen, and that they would just be listening closely to the target sentences read aloud by a single speaker. The post-test session was exactly the same for the control subjects and the trained subjects. Twenty-one trials were presented, including one attention-check trial set to occur halfway through the session. All target sentences presented during the post-test session were novel. No breaks were offered during the post-test session.

Demographic and language background questionnaire. After a reminder that responses would not affect pay for the study, a series of questions asked subjects to report their age, gender, race, ethnicity, hearing status, nationality, and languages. Lastly, subjects were asked to report if they thought that there was any reason that their data should be excluded from the study.

Data preparation. Transcription accuracy for both the training and the post-test sessions was calculated with the R package *autoscore* (Borrie et al., 2019). Before using autoscore, keywords in the Hearing in Noise Test sentences were identified for each sentence. Specifically, determiners were excluded from scoring (e.g., “the”, “a”, “an”, “his”, “her”). The tallied number of correct versus incorrect (missed) keywords per sentence was used for analyses. In the training session data, the effect of trial was renumbered from 0 to 59 (i.e., differences at the intercept represent differences at Trial 1), and then scaled to assist with model convergence.

4.1.2 Results

Analyses of the training data are reported first, followed by analyses of the post-test data. For both datasets, we used the *glmer()* function from the *lme4* package in R to fit the data from the speech transcription task. The predicted variable, transcription accuracy, was treated as a grouped binomial. That is, each trial of the task (for a given subject) corresponded to a single row of the dataframe in R, and the predicted variable of the models was two columns: correctly identified target words and missed target words. A logit link function was specified for the models. Random effects included intercepts for subjects and items.

Training results. Fixed effects in the model included: prime (levels: East Asian, White; dummy-coded reference level: East Asian), trial, and the interaction between prime and trial.

Log-likelihood model comparisons were used to determine whether each fixed effect significantly improved model fit. The effect of prime did not significantly improve model fit ($\chi^2(1) = 0.66$, $p = .42$; **Figure 4.1**). The effect of trial did significantly improve fit ($\chi^2(1) = 111.18$, $p < .001$), indicating that subjects improved across the session, but the interaction between prime and trial was not significant ($\chi^2(1) = 2.02$, $p = .16$), nor were differences between prime at the intercept.

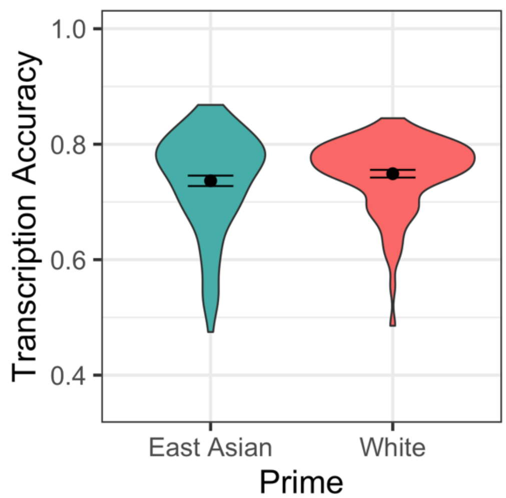

Figure 4.1. By-subject transcription accuracy during the training session is summarized using proportion of words correctly transcribed. Violin plots show distribution of subjects' average performance and points show group means with standard error bars.

Post-test results. For the post-test dataset, the only fixed effect examined was condition (levels: control group, East Asian primed group, White primed group). Our initial dummy-coding scheme set the control condition as the reference level, and in a separate model we changed the reference level to East Asian primed group (to directly compare the two training groups). A log-likelihood model comparison indicated that the effect of condition significantly improved model fit ($\chi^2(2) = 10.38$, $p = .006$). When the control condition was set as the reference level, model estimates indicated that subjects assigned to the White prime training performed significantly better than those in the control group ($\beta = 0.27$, $p = .001$), and those assigned to the

East Asian prime training performed marginally better ($\beta = 0.15$, $p = .08$; **Figure 4.2**). When the East Asian prime condition was set as the reference level, the model estimate for the difference between the East Asian and White prime training groups was not significant ($\beta = 0.13$, $p = .13$).

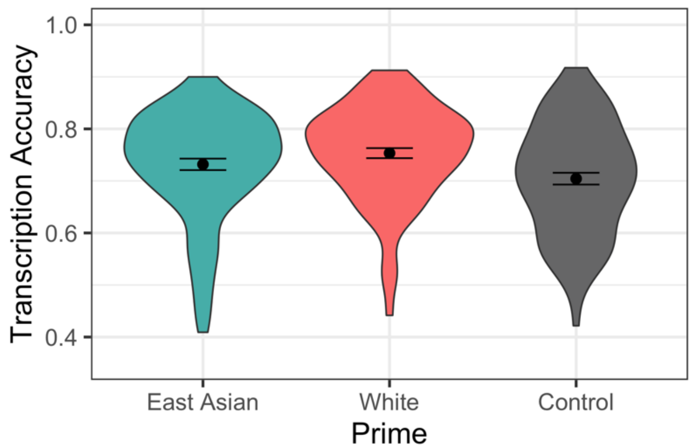

Figure 4.2. By-subject transcription accuracy during the post-test session is summarized using proportion of words correctly transcribed. Violin plots show distribution of subjects' average performance and points show group means with standard error bars.

4.1.3 Discussion

In Experiment 4, the effect of social priming on perceptual adaptation was examined in a multi-talker training session with a novel-talker post-test. Subjects trained with East Asian prime images versus subjects trained with White prime images did not perform significantly differently during the training session. Additionally, these two groups did not differ in their performance at post-test. Together, these results indicate that our matched-guise manipulation of speaker race did not affect perceptual adaptation to the Cantonese-accented talkers, or the generalizability of perceptual learning to a novel talker of the same nonnative accent.

In line with prior work (Bradlow & Bent, 2008), subjects given a training session performed better at the post-test session than subjects given no training (i.e., the control group). This outcome indicates that listeners were accommodating the Cantonese-accented English speech during the training session, and then generalized this perceptual learning when they encountered a novel Cantonese-accented talker. Surprisingly, subjects assigned to the White prime training session performed significantly better than the control group, while subjects assigned to the East Asian prime training session only performed marginally better than the control group. Although the two priming groups did not differ when directly compared, these trends do indicate that the White prime training may have better prepared subjects for the post-test session. This trend is the opposite of what we predicted based on prior work that has found benefits of East Asian primes for the perception of Mandarin Chinese-accented English (McGowan, 2015; although see mixed evidence in Experiments 1 and 2).

One notable difference between the design of Experiment 4 and the design of Experiments 1-3 was the use of speech-shaped noise in place of informational (multi-talker babble) noise. For Experiment 4, we determined that speech-shaped noise was better-suited for the multi-talker training session (i.e., it is especially difficult to identify a target talker among other voices when the target talker changes each trial). However, the effect of social primes on transcription accuracy for nonnative accent has yet to be compared across speech-in-noise settings. It is possible that socio-indexical cues are more greatly utilized in listening environments with multiple talkers. If this is the case, it would indicate a much more situationally-dependent effect than we had originally hypothesized.

5. General Discussion

This dissertation investigated the effect of speaker race/ethnicity information on the perception of native- and nonnative-accented speech. Here, we integrate results from multiple experiments to draw conclusions on the following topics: 1) The generalizability and specificity of social priming effects; 2) Whether there is a link between implicit associations and the effects of social primes on speech perception; and, 3) Whether social primes can benefit the listener during perceptual adaptation to unfamiliar, nonnative-accented speech. Limitations of the current studies and future directions for examining social priming of native- and nonnative-accented speech are also discussed.

5.1 Generalizability of social priming effects

Previous work examining social priming has focused on the effects of White and East Asian face primes on perception of native- and Mandarin Chinese-accented English. To examine the generalizability of social priming effects, we began by extending this line of inquiry to a Middle Eastern prime and Arabic-accented English. We predicted that subjects assigned to view a Middle Eastern woman’s face during perception of Arabic accent would outperform subjects assigned to view a White woman’s face or a control image. However, results indicated no difference in performance between the subjects assigned to the White, Middle Eastern, or control prime.

For our next test of the generalizability of social priming effects, we compared the effects of a White and a Latina face prime on the perception of American-accented English speech. In this case, we predicted better performance by subjects assigned to view a White prime. This experiment was most comparable to Babel and Russell (2015), in which Chinese face primes resulted in poorer transcription accuracy for Canadian-accented English (as compared to a fixation cross). To our surprise, there was once again no effect of the primes on speech transcription accuracy.

Together, the evidence from these two experiments do not indicate that social priming effects generalize to additional races/ethnicities and/or nonnative accents. However, there were also multiple instances in which our White and East Asian priming experiments failed to replicate prior work. Our initial replication of McGowan (2015) was promising: Results indicated that subjects randomly assigned to view an East Asian prime significantly outperformed subjects assigned to view a White prime when transcribing sentences produced by a Mandarin-accented talker. However, after we expanded the size of our dataset (from $N = 50$ to $N = 80$ per group) this difference was no longer significant. A similar null result was found in our final experiment, in which we examined the effect of social priming on perceptual adaptation. Subjects randomly assigned to view East Asian faces during a Cantonese-accent training did not perform significantly differently from subjects assigned to view White faces. Indeed, even data examining the effects of an East Asian and White prime on American-accented English (similar to Babel & Russell, 2015) indicated no difference between conditions.

Thus, the strongest evidence found for an effect of race/ethnicity primes on speech recognition accuracy was our conceptual replication of McGowan (2015), and this significant outcome was outnumbered by the null results of our subsequent experiments. Regarding the question of generalizability, it is difficult to draw conclusions from the current state of evidence. Despite our use of large sample sizes across all experiments, we did not find evidence to indicate a generalization of social priming effects to additional races/ethnicities. However, given the null results for our experiments with White and East Asian primes, it is possible that this is attributable to differences between the population we sampled and the populations sampled in prior work. We discuss this possibility in Section 5.6.

5.2 Specificity of social priming effects

To examine the specificity of social priming effects, we expanded on our replication of McGowan (2015). Two key conditions were included in the experiment to examine this aim: A Middle Eastern prime combined with Mandarin accent, and an East Asian prime combined with Arabic accent. We hypothesized that social priming had either a high degree of specificity (such that listeners associate specific races/ethnicities with specific accents) or a low degree of specificity (such that listeners broadly associate minority races/ethnicities with nonnative accents). The former of these hypotheses was aligned with an exemplar theory account of social priming: Episodic traces stored in the lexicon link linguistic aspects of a specific accent (e.g., Mandarin Chinese) with social qualities of the speaker (e.g., East Asian), such that visual cues to race/ethnicity can activate relevant phonemic categories.

Results indicated that subjects assigned to a Middle Eastern prime when perceiving Mandarin accent performed no differently than subjects assigned to an East Asian, White, or control prime. However, trends in the data did indicate that subjects in the Middle Eastern prime condition performed more similarly to subjects in the East Asian prime conditions (i.e., both groups trended toward better performance than the White prime group). For the Arabic accent conditions, a different trend emerged. Subjects assigned to the East Asian prime performed significantly worse than those assigned to the control prime, and trends in the data indicated relatively (but not significantly) poorer performance as compared to the White and Middle Eastern primes as well. Thus, the outcomes of the Mandarin-Middle Eastern and the Arabic-East Asian conditions appear to be at odds. Trends in the Mandarin accent data appear to support the low-specificity hypothesis, while trends in the Arabic accent data appear to support the high-specificity hypothesis.

Given the inconsistency in these results and the marginally-significant trends, this research question will need to be revisited in future research. However, our data do provide an important insight: Specifically, it is possible that social priming effects may not be uniformly high or low specificity. Some races/ethnicities may be associated with specific accents, while others may not be. In other words, the listener's degree of familiarity with a specific accent may modulate social priming effects. For the American subjects in the current study, it is possible that they were more familiar with East Asian nonnative accents than Middle Eastern nonnative accents. This would account for why the East Asian-Arabic group showed markedly poorer performance, while the Middle Eastern-Mandarin group showed relatively better performance. In future work, a more in-depth analysis of subjects' exposure to relevant accent varieties and the racial/ethnic diversity of their social networks could allow us to address this hypothesis.

5.3 Implicit associations and social priming

One of our primary aims was to improve our understanding of the mechanism supporting social priming during speech perception. The prior state of evidence indicated that social priming may be best captured with an exemplar model of speech perception, in which the lexicon contains episodic representations of words and phonemes. Based on an exemplar model account, we thus hypothesized that socio-phonetic links should also vary across listeners based on their unique experiences. In this case, we would thus expect that individuals with stronger associations between given racial and accent categories would have larger social priming effects. We aimed to test this using Implicit Association Tests (IATs) as measures of individual differences in listeners' implicit racial/ethnic associations.

Two IATs were included in our examination of social priming for (native) American-accented English speech. The first of these IATs examined associations between the two races/ethnicities of interest (White and East Asian, or White and Latinx) and the constructs *American* vs. *Foreign*. We predicted that this IAT would be most likely to capture individual differences relevant for the social priming experiment (i.e., that it would capture associations with American vs. foreign accents). The second IAT examined associations between the races/ethnicities of interest and the constructs *Good* vs. *Bad*. We predicted that scores from the Good-Bad IAT would not predict social priming effects.

For both cohorts in the study, neither type of IAT scores significantly interacted with social priming effects. Thus, our evidence does not indicate that IATs capture implicit racial/ethnic associations related to the social priming of speech. There were trends in the data, however, that in some cases indicated a relationship between IAT scores and overall performance on the speech perception task. Specifically, subjects who had stronger implicit associations in the predicted direction (e.g., White + American and East Asian + Foreign) also had lower overall speech recognition scores. We suggest that this correlation between IAT scores and performance on the speech perception task may be explained by a third variable. Indeed, individual differences in executive cognitive control (e.g., inhibition, working memory, and task switching) can predict IAT scores (Klauer et al., 2010; Ito et al., 2015), such that individuals with greater executive function skills tend to have smaller IAT scores (because they are better able to control their responses during the IAT). For perception of speech-in-noise, individual differences in executive control (particularly working memory; Zekveld et al., 2013; Yeend et al., 2019) can also predict recognition accuracy. Thus, the negative correlation between the IAT and the speech transcription task may actually be supported by a third-variable relationship with executive cognitive control.

While the results of our study did not indicate that the IATs capture meaningful variation in social priming effects, it is important to note that we also failed to find a significant overall effect of social priming. It is possible that our ability to examine the relationship between implicit racial/ethnic associations and social priming was hindered by this null outcome. Additionally, the selection of an American-Foreign IAT relied on an assumption that the constructs American and Foreign captured native- vs. foreign-accentedness. The benefit of using the American-Foreign IAT was that it had been used in prior work. However, for future research, it may be beneficial to develop and validate an IAT for direct examination of implicit associations between race/ethnicity and accentedness (e.g., Pantos & Perkins, 2013).

5.4 Social priming in the context of perceptual adaptation

In our replication of McGowan (2015), we incorporated a novel approach to analyzing the data. By including trial number in our analysis, we were able to estimate the rate at which listeners in each priming condition perceptually adapted to the Mandarin accent. Subjects randomly assigned to view an East Asian face had an immediate benefit as compared to those assigned to view a White face, indicating a rapid and automatic effect of the primes on speech perception. Additionally, subjects assigned to the control condition showed a significantly faster rate of adaptation than either of the other subject groups. Visually, the trends in the data indicated that subjects in the control and White prime groups began at a similar level of performance in Trial 1, but that the control group was uniquely unhindered. By the final trial of the experiment, subjects assigned to the control prime reached a similar level of performance as subjects assigned to the

East Asian prime. By incorporating an analysis of perceptual adaptation to the Mandarin accent, our understanding of the effect of the social primes on speech perception was greatly improved.

However, the results of our subsequent experiment, which increased the sample size of the McGowan (2015) replication and added additional conditions, complicated our initial conclusions. The differences found in our replication study between the East Asian and White prime conditions were no longer significant after increasing our sample size. Additionally, our examination of perceptual adaptation for an Arabic accent did not reveal any immediate benefits (i.e., differences at the intercept) of the Middle Eastern prime over the control or White primes. The one thing that remained consistent across experiments was the significantly faster rate of improvement by control group subjects. For both the Mandarin and Arabic accent conditions, subjects randomly assigned to view a silhouette control face with no race/ethnicity information improved the most across the experiment. One possible explanation for this outcome is that subjects in the control condition are gaining more from the acoustic signal overall. Without a prime, control subjects may begin at a relative disadvantage, but they are also able to gather similar socio-phonetic information over time. Additionally, the lack of a priming image may encourage control subjects to rely more heavily on the acoustic signal when adapting to the nonnative accent.

In a separate study, we also examined multi-talker perceptual adaptation to Cantonese accent, and generalization of perceptual learning to a novel speaker of the same nonnative accent. We hypothesized that subjects assigned to a training session with East Asian primes would outperform subjects assigned to a session with White primes, both during training and at post-test. The effect of the priming manipulation was non-significant in both the training and the post-test data. At post-test, a control group that received no training whatsoever was included, and results did indicate that subjects from the trained groups outperformed these controls. This comparison confirmed that exposure to Cantonese accent prior to the post-test provided a benefit (as would be expected based on prior work; Bradlow & Bent, 2008). The results of this experiment, in combination with the null results of the other social priming experiments in this dissertation, indicate that the effect of social primes may be less robust than previously assumed.

5.5 Limitations of the current work and future directions

The most notable difference between the current dissertation and prior work examining social priming effects (e.g., McGowan, 2015; Babel & Russell, 2015) was the method of participant recruitment. Due to the COVID-19 pandemic, all of the experiments were designed for online, remote participation. A common concern with online data collection is that the researcher has less control over the experiment than they otherwise would for in-person data collection. Thus, to increase the quality of our data, we took multiple measures to ensure participant engagement and audio quality: 1) All experiments contained attention-check trials to ensure subjects were not making random-string entries; 2) Subjects were given a non-penalized opportunity to report if there was any reason to exclude their data (e.g., "I was too distracted during the experiment"); and, 3) Subjects were given a non-penalized opportunity to report if they failed to use headphones, and in some cases an additional headphone screening task was implemented. Lastly, because two of our experiments recruited university students, and two recruited online paid subjects, it is unlikely that the null results seen across these experiments is due to the use of online paid subjects.

Thus, we do not have reason to suspect that conducting these same experiments in-lab would change the outcomes. However, we do note that by sampling subjects from across the United States, we may have inadvertently introduced noise into our sample. Some communities may have larger nonnative speaker populations, and others smaller, resulting in different socio-phonetic mappings in the native listeners. For example, Babel and Russell (2015) sampled subjects from the Metro Vancouver area, where stereotypes of Asian Canadians as second language speakers of English are particularly prominent. In future work, including measures of exposure to relevant nonnative accents and of the diversity of individuals' social networks may help to clarify the role of listener experience in social priming effects.

Another possible limitation of the current work is the use of transcription accuracy in the speech perception tasks. To avoid ceiling effects, all of the auditory stimuli were mixed with background noise, which may have reduced listeners' abilities to identify and/or adapt to the speakers and their accents. The addition of background noise, while necessary for transcription accuracy-based tasks, could be avoided with another psycholinguistic method (e.g., lexical decision tasks, dual-task paradigms, pupillometry, or eye-tracking). Additionally, it may be beneficial in future work to use phoneme or syllable-length materials (i.e., in place of sentence-length materials), which can allow the researcher to highlight specific attributes of an accent. Ultimately, by incorporating additional methodological approaches, we will be able to determine whether social priming effects are merely small (i.e., resulting in the numerous null outcomes of the present study), or whether they may be context-dependent (e.g., less easily observed in speech-in-noise).

5.6 Conclusion

A growing body of evidence indicates that listeners exploit socio-indexical information to identify social characteristics of the speaker and accommodate their accent. One small portion of this literature has examined the effect of race/ethnicity information, manipulated via visual guises, on speech transcription accuracy for sentence-length materials presented in noise. In this dissertation, we aimed to build on this work. Our results indicate that social priming effects for native- and nonnative-accented speech may either be smaller or more context-dependent than previously assumed. Indeed, we found evidence of race/ethnicity primes affecting speech transcription accuracy and/or perceptual adaptation in only one of four experiments. Importantly, our null outcomes have one positive implication: Minority race/ethnicity speakers with native-accented English speech may not be at a perceptual disadvantage as compared to White speakers (as indicated by prior work). However, given the mixed outcomes of our study, we recommend further examination of social priming effects for native- and nonnative-accented speech. Additionally, we suggest that by including in-depth analyses of subjects' social networks (i.e., to determine their exposure to racial/ethnic diversity and relevant accent varieties), future investigations may be able to assess individual differences in susceptibility to social priming effects.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407.

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. The Journal of the *Acoustical Society of America, 137*(5), 2823–2833.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174-EL180.

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Audiovisual cues benefit recognition of accented speech in noise but not perceptual adaptation. *Frontiers in Human Neuroscience, 9*, 422.

Bent, T., Holt, R. F., Miller, K., & Libersky, E. (2019). Sentence Context Facilitation for Children’s and Adults' Recognition of Native- and Nonnative-Accented Speech. *Journal of Speech, Language, and Hearing Research: JSLHR, 62*(2), 423–433.

Bieber, R. E., & Gordon-Salant, S. (2017). Adaptation to novel foreign-accented speech and retention of benefit following training: Influence of aging and hearing loss. *The Journal of the Acoustical Society of America, 141*(4), 2800-2811.

Boersma, Paul & Weenink, David (2021). *Praat: doing phonetics by computer* [Computer program]. Version 6.1.10, retrieved 1 July 2017 from http://www.praat.org/

Borrie, S. A., Barrett, T. S., & Yoho, S. E. (2019). Autoscore: An open-source automated tool for scoring listener perception of speech. *The Journal of the Acoustical Society of America, 145*(1), 392.

Bradlow, A. R. (n.d.) *SpeechBox.* Retrieved from https://speechbox.linguistics.northwestern.edu

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339–2349.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729.

Bradlow, A. R., Blasingame, M., & Lee, K. (2018). Language-independent talker-specificity in bilingual speech intelligibility: Individual traits persist across first-language and second-language speech. *Laboratory Phonology, 9*(1). https://doi.org/10.5334/labphon.137

Brennan, E. M., & Brennan, J. S. (1981). Accent Scaling and Language Attitudes: Reactions to Mexican American English Speech. *Language and Speech, 24*(3), 207–221.

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Morinaga Shearman, S. (2002). Attitudinal and affective response toward accented English. *Language & Communication, 22*(2), 171–185.

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. (2020). Author accepted manuscript: Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology*, 1747021820916726.

Campbell-Kibler, K. (2012). The Implicit Association Test and sociolinguistic meaning. Lingua. International Review of General Linguistics. *Revue Internationale de Linguistique Generale, 122*(7), 753–763.

Carlson, H. K., & McHenry, M. A. (2006). Effect of accent and dialect on employability. *Journal of Employment Counseling, 43*(2), 70–83.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647–3658.

Craik, F. I., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *The Quarterly Journal of Experimental Psychology, 26*(2), 274-284.

De Meo, A. (2012). How credible is a non-native speaker? Prosody and surroundings. *Methodological Perspectives on Second Language Prosody Papers from ML2P*, *2012*, 3-9.

Devos, T., & Banaji, M. R. (2005). American= white? *Journal of Personality and Social Psychology, 88*(3), 447.

Dragojevic, M., & Giles, H. (2014). The reference frame effect: An intergroup perspective on language attitudes. *Human Communication Research*. Retrieved from https://academic.oup.com/hcr/article-abstract/40/1/91/4093728

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*. Retrieved from https://academic.oup.com/hcr/article-abstract/42/3/396/4064730

Dragojevic, M., Giles, H., Beck, A.-C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs, 84*(3), 385–405.

Elman, J. L., & McClelland, J. L. (1983). Exploiting lawful variability in the signal: The TRACE model of speech perception. *The Journal of the Acoustical Society of America, 74*(S1), S68-S68.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology, 42*(1), 120–133.

Getty Images. (2020, December). *iStock.* https://www.istockphoto.com/

Gluszek, A., & Dovidio, J. F. (2010). Speaking With a Nonnative Accent: Perceptions of Bias, Communication Difficulties, and Belonging in the United States. *Journal of Language and Social Psychology, 29*(2), 224–234.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition, 22*(5), 1166.

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

Green, P., & MacLeod, C. J. (2016). SIMR : an R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution / British *Ecological Society, 7*(4), 493–498.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*. Retrieved from https://www.degruyter.com/view/j/tlir.2006.23.issue-3/tlr.2006.014/tlr.2006.014.xml

Heaton, H., & Nygaard, L. C. (2011). Charm or Harm: Effect of Passage Content on Listener Attitudes toward American English Accents. *Journal of Language and Social Psychology, 30*(2), 202–211.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*(2), 96-101.

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology, 108*(2), 187–218.

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 19*(6), 741–758.

Johnson, K. (1997). *Speech perception without speaker normalization: an exemplar model*. In Talker Variability in Speech Processing, K. Johnson and J. Mullenix (eds.), 145-166. Academic Press.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*(4), 359–384.

Kang, O., & Rubin, D. L. (2009). Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation. *Journal of Language and Social Psychology, 28*(4), 441–456.

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the implicit association test: why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology, 63*(3), 595–619.

Kraus, M. W., Torrez, B., Park, J. W., & Ghayebi, F. (2019). Evidence for the reproduction of social class in brief speech. *Proceedings of the National Academy of Sciences of the United States of America, 116*(46), 22998–23003.

Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *The Journal of the Acoustical Society of America, 29*(1), 98–104.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review, 74*(6), 431.

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States.* Routledge.

Lockhart, R. S. (2002). Levels of processing, transfer-appropriate processing, and the concept of robust encoding. *Memory, 10*(5-6), 397-403.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135.

Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R., & Saltzman, D. (2021). Talker familiarity and the accommodation of talker variability. *Attention, Perception, & Psychophysics, 83*(4), 1842-1860.

McGowan, K. B. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech, 58*(Pt 4), 502–521.

McGowan, K. B., & Babel, A. M. (2019). Perceiving isn't believing: Divergence in levels of sociolinguistic awareness. *Language In Society*, 1–26.

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception & Psychophysics, 80*(6), 1559–1570.

McLaughlin, D. J., Brown, V. A., Carraturo, S., & Van Engen, K. J. (2022). Revisiting the relationship between implicit racial bias and audiovisual benefit for nonnative-accented speech. *Attention, Perception & Psychophysics*. https://doi.org/10.3758/s13414-021-02423-w

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2020). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, 1-12.

Munro, M. J., & Derwing, T. M. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning, 45*(1), 73–97.

Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics, 35*(1), 85-103.

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology, 18*(1), 62–85.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085–1099.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology, 47*(2), 204-238.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality & Social Psychology Bulletin, 31*(2), 166–180.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test. *Experimental psychology*, *54*(1), 14-29.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 19*(2), 309.

Pantos, A. J., & Perkins, A. W. (2013). Measuring Implicit and Explicit Attitudes Toward Foreign Accented Speech. *Journal of Language and Social Psychology, 32*(1), 3–20.

Pierrehumbert, J. (2001). Lenition and contrast. *Frequency and the Emergence of Linguistic Structure, 45*, 137.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and Phonetic Experiments on American English Dialect Identification. *Journal of Language and Social Psychology, 18*(1), 10–30.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511–531.

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment, 24*(4), 210.

Strand, E. A. (1999). Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology, 18*(1), 86–100.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience, 8*, 577.

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.

Wilson, T. D., & Dunn, E. W. (2004). Self-knowledge: its limits, value, and potential for improvement. *Annual Review of Psychology, 55*, 493–518.

Xie, X., Theodore, R. M., & Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 43*(1), 206.

Xu, F. K., Nosek, B. A., Greenwald, A. G., Ratliff, K., Bar-Anan, Y., Umansky, E., Banaji, M. R., Lofaro, N., Smith, C., & Axt, J. (2013, September 6). *Project implicit demo website datasets.* OSF. Retrieved January 4, 2022, from https://osf.io/y9hiq/

Yeend, I., Beach, E. F., & Sharma, M. (2019). Working Memory and Extended High-Frequency Hearing in Adults: Diagnostic Predictors of Speech-in-Noise Perception. *Ear and Hearing, 40*(3), 458–467.

Yi, H.-G., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America, 134*(5), EL387–93.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America, 134*(3), 2225–2234.

Zheng, Y., & Samuel, A. G. (2020). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(7), 1270.

Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of Personality and Social Psychology, 112*(5), 696–717.