Social Priming of Speech Perception: The Role of Individual Differences in Implicit Racial and Ethnic Associations

Drew J. McLaughlin[1,2]

Kristin J. Van Engen[1]

[1] Department of Psychological & Brain Sciences, Washington University in St. Louis

[2] Basque Center on Cognition, Brain and Language

Author Note

Acknowledgements: This work was supported by National Science Foundation Graduate Research Fellowship DGE-1745038 (Drew J. McLaughlin), a National Science Foundation Doctoral Dissertation Research Improvement Grant 2116319 (Kristin J. Van Engen and Drew J. McLaughlin), and by the Basque Government through the BERC 2022-2025 program and the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S.

The authors have no relevant financial or non-financial conflicts of interest to disclose.

Data and code are available from: https://osf.io/nd7wm/files/osfstorage

This study's design was pre-registered, see: https://osf.io/vdazs; https://osf.io/36v9x

Correspondence concerning this article should be addressed to Drew J. McLaughlin, Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 20009 Donostia-San Sebastián, Gipuzkoa, Spain. Email: d.mclaughlin@bcbl.eu. Phone: +34 943 309 300

Word count: 23,103

Abstract

Prior research has shown that visual information, such as a speaker's perceived race or ethnicity, prompts listeners to expect a specific socio-phonetic pattern ("social priming"). Indeed, a picture of an East Asian face may facilitate perception of second language (L2) Mandarin Chinese-accented English but interfere with perception of first language- (L1-) accented English. The present study builds on this line of inquiry, addressing the relationship between social priming effects and implicit racial/ethnic associations for L1- and L2-accented speech. For L1-accented speech, we found no priming effects when comparing White versus East Asian or Latina primes. For L2- (Mandarin Chinese-) accented speech, however, transcription accuracy was slightly better following an East Asian prime than a White prime. Across all experiments, a relationship between performance and individual differences in implicit associations emerged, but in no cases did this relationship interact with the priming manipulation. Ultimately, exploring social priming effects with additional methodological approaches, and in different populations of listeners, will help to determine whether these effects operate differently in the context of L1- and L2-accented speech.

Keywords: speech perception, social priming, implicit bias, language attitudes

Public significance statement:

This study suggests that a speaker's race may impact how well we are able to understand and transcribe a foreign accent, but not necessarily how well we are able to understand and transcribe a native accent. A listener's implicit racial and ethnic biases do not appear to affect how well they are able to understand either foreign or native accented speech.

Social Priming of Speech Perception: The Role of Individual Differences in Implicit Racial and Ethnic Associations

An abundance of social information can be conveyed by a person's speech, such as their age, race, gender, sexual orientation, and social class (Strand, 1999; Labov, 1986; Munson & Babel, 2007). Additionally, a regional or second-language (L2) accent can signal to the listener whether a speaker belongs to their "in-group" or an "out-group" (Lippi-Green, 2012). Indeed, language users often use the term "accented" as a way of expressing that a given speaker is different than themselves. A talker from New York may call a talker from Texas "accented," and a talker from Texas may conversely call a talker from New York "accented." All speech has an accent, but, colloquially, the term "accented" is used by language users to categorize other talkers, and these categorizations can be linked to expectations about social class and other speaker traits.

As our world becomes more globalized, listeners are increasingly likely to encounter L2 ("foreign"[1]) accents in day-to-day life. Perceiving spoken language requires listeners to map complex acoustic input onto linguistic representations in their mental lexicons – a process that can be complicated by unfamiliar L2 accent. L2-accented speech is characterized by systematic and idiosyncratic segmental and suprasegmental differences from first language (L1) productions (e.g., Wang & van Heuven, 2006). Thus, matching L2-accented productions to internal perceptual categories based on L1 productions can be a cognitively demanding and time-consuming task for listeners (Van Engen & Peelle, 2014; McLaughlin & Van Engen, 2020; Brown, McLaughlin, Strand, & Van Engen, 2020; Floccia, Butler, Goslin, & Ellis, 2009; Clarke

[1] We use the term "foreign" accent here because it is the colloquial term most often used to refer to L2 accents. We note, however, that this term is not exclusively used to refer to L2 accents, but may also be used to refer to L1 varieties from regions that are less linguistically prestigious (e.g., Indian English).

& Garrett, 2004), and may ultimately result in poorer speech recognition (or “intelligibility”). Although listeners also appear to be able to rapidly adapt to unfamiliar L2 accents (Clarke & Garrett, 2004; Brown et al., 2020), and familiarity with an L2 accent reduces the cognitive demands of speech processing (Porretta & Tucker, 2019).

From a very young age, humans begin to develop attitudes and preferences about race, language, accent, and nationality (Kinzler, Dupoux, & Spelke, 2007; Kinzler, Shutts, DeJesus, & Spelke, 2009; Kinzler & Spelke, 2011; DeJesus, Hwang, Dautel, & Kinzler, 2018). Prior research has shown that social cues from *outside* the acoustic signal can even affect speech processing (i.e., via expectations of a given accent or vocal qualities). For example, listeners categorize fricatives in an /s/ to /ʃ/ continuum differently depending on whether they perceive a talker to be a man or a woman (manipulated experimentally; Strand, 1999); inferences about the talker’s gender can be drawn from both the adjacent vowel to /s/ or /ʃ/ (a speech-internal cue), or from the image of the speaker’s face (a speech-external and non-linguistic cue). 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 with the speech perception process. In the present study, we aimed to build on the social priming literature and examine the role of listeners’ implicit racial and ethnic associations.

Social priming

While the current study will focus on the effects of visual race and ethnicity guises on speech recognition, many notable social priming studies have examined the effects of verbal (i.e., written) guises on speech recognition. For example, 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. In the study, subjects were presented with sentences and told to concentrate on a key vowel; next, they were instructed to match this vowel to one from a set of six computer-resynthesized options. Most notably, even though the same stimulus (a raised /a/ produced by a Detroit speaker) was presented to each group, subjects given the "from Ontario, Canada" guise were more likely to match it to the raised synthesized /a/ vowel while subjects given the "from Detroit, USA" guise were more likely to match it to the canonical or low synthesized /a/ vowel. 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; cf, Walker et al., 2019; see also McGowan & Babel, 2019).[2]

Using a visual matched-guise paradigm, Rubin (1992) found that American listeners had poorer comprehension of a short lecture when shown a picture of an East Asian face than when shown a White face. In both conditions, the lecture recording was the same L1 American-accented speaker. Thus, Rubin interpreted this outcome as an effect of listeners' biases on speech comprehension. These results were later replicated in a study by Kang and Rubin (2009), which also extended the prior work by examining listeners' stereotype judgments under each guise condition. Most notably, when shown an East Asian guise, subjects rated the speaker as sounding more foreign-accented.

Babel and Russell (2015) and McGowan (2015) were the first studies to examine the direct effect of social priming on speech recognition accuracy for L1 English listeners. With a sample of subjects from Vancouver, BC, Canada, Babel and Russell examined social priming effects for L1 English listeners when presented with L1 English speech in pink noise. Their

[2] It is worth noting that these studies involving vowel-matching differ methodologically from the other speech perception studies highlighted in this section. The nature of the stimulus-to-vowel matching process requires a larger memory-encoding and retrieval aspect.

findings indicated that listeners' ability to recognize L1 speech produced by Chinese-Canadian talkers was reduced when the talkers' faces were presented on screen (as compared to when a fixation cross was presented on screen). For White-Canadian talkers, however, this effect did not occur. Thus, it appeared that expectations about Chinese-Canadians' accents negatively affected speech perception – even though the speakers had L1-accented speech. Complementing the design of Babel and Russell (2015), McGowan (2015) examined social priming effects for L1 English listeners when presented with L2, Mandarin Chinese-accented English speech. In this study, McGowan found that American listeners had better recognition accuracy when presented with an East Asian face than a White face. Together, these results indicate that the outcome of Babel and Russell (2015) may reflect an automatic social priming cost. Indeed, Babel and Russell suggest that the faces of the Chinese-Canadian speakers presented in their experiment may have activated socio-phonetic categories for L2, Chinese-accented English; thus, when listeners encountered L1, Canadian-accented English speech, this perceived incongruency may have hindered speech recognition accuracy.

Building on this line of inquiry, recent work by Kutlu, Tiv, Wulff, and Titone (2022) found that, for perception of both Indian and British L1 English accents, American listeners (from Gainesville, Florida) and Canadian listeners (from Montreal, Canada) alike had poorer recognition accuracy when presented with South Asian faces as compared to White faces – although this negative priming effect was more prominent in the American listeners. The authors also examined recognition accuracy for American-accented English and did not find any effects of the priming manipulation. As part of this same study, Kutlu and colleagues also examined accent ratings, and found that the American listeners rated the American and British English accents as sounding more accented when paired with the South Asian faces as compared to the

White faces. Thus, the results of their study demonstrated that listeners with different language experiences (e.g., a predominantly multi-lingual versus mono-lingual upbringing) may be differentially affected by social cues, such as a speaker's race. It is worth noting, however, that work by Zheng and Samuel (2017) has demonstrated that accentedness ratings can be affected by demand characteristics. In other words, subjects appear to change their behavior based on what they believe the researcher's hypothesis is. For accentedness ratings following social primes, there is a possibility that subjects may be rating the minority speaker as sounding "more accented" because they assume that is what the experimenter aims to discover in their research. In the present study's review of the social priming literature, we focus on prior findings for speech comprehension and intelligibility (i.e., measures of performance). We interpret these as measures of perception, as opposed to self-ratings, and therefore less susceptible to demand characteristics.

Matching the results of McGowan (2015), in teens and older adult German L1 listeners, Hanulíková (2021) found that Korean-accented German speech was more accurately perceived when presented with an East Asian prime than a White prime. For a group of young adult German listeners also included in the experiment, however, this effect was not significant. Further, for a dominant ("Standard German") L1 accent, no effects of priming were found for any group, and for a less common ("Palatinate German") L1 accent, no significant effects emerged (although trends favored the White prime condition over the East Asian prime condition). The results of Hanulíková (2021) thus extend findings such as McGowan (2015), but also indicate that social priming effects may vary across age groups – possibly reflecting different accrued experiences with race and accents.

Further work has also re-examined social priming for L2 accent, with varying outcomes. In a recent conceptual replication and expansion of McGowan (2015), our lab found that White American listeners were better able to understand Mandarin-accented English when paired with an East Asian, as compared to a White, face (McLaughlin & Van Engen, under review). Our data further showed that this difference was significant beginning at Trial 1, demonstrating that the priming effect was extremely rapid. Notably, however, the difference between these priming conditions was not significant in a follow-up experiment with a larger sample size. Further, when examining social priming for Arabic-accented English (paired with a silhouette, White, Middle Eastern, or East Asian prime) no significant social priming effects were found. Similarly, Melguy and Johnson (2021) examined the effects of a silhouette, European (White), East Asian, and South Asian prime on perception of Mandarin-accented English for American listeners and found no differences in transcription accuracy between conditions. The authors did, however, find that subjects who reported that the Mandarin-accented speaker sounded L2-accented significantly outperformed those who reported that the speaker sounded L1-accented – even when the correct L2 accent (Mandarin Chinese) was not identified.

A Framework for Social Priming

Exemplar theory proposes that episodic traces are encoded in the lexicon (Hintzman, 1984; Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001; Johnson, 2006). Many researchers have suggested that non-auditory factors, such as characteristics of the speaker, are also stored with these exemplars (see work on talker familiarity effects: Craik & Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993; Newman & Evers, 2007; Magnuson, Nusbaum, Akahane-Yamada, & Saltzman, 2021). Over time, listeners may create abstracted categories, systematically linking social groupings and phonetic patterns (as proposed by Melguy & Johnson, 2021). On such a view, non-linguistic information such as a speaker's perceived gender (Strand, 1999) or race/ethnicity (Babel & Russell, 2015) would be able to activate exemplars stored in the mental lexicon, causing top-down speech processing effects (Johnson, Strand, & D'Imperio, 1999). Additionally, the efficiency of speech recognition ought to be influenced by the number of similar exemplars in memory; thus, processing speech produced by a familiar talker (Newman & Evers, 2007; Magnuson, Nusbaum, Akahane-Yamada, & Saltzman, 2021) or in a familiar accent should be faster. More generally, based on an exemplar model, one would expect social priming effects to vary across listeners based on their unique experiences. For example, individuals with stronger associations between given racial/ethnic and accent categories should have larger social priming effects, because these connections between social cues and linguistic categories ought to be more robust; here, the quality and detail of the linguistic representations may also matter, with richer representations resulting in more "robust" connections. For example, if a Chinese face is presented to a listener with a weak (or no) association between Chinese faces and foreign accents, then we would expect this listener to have a very small (or no) benefit of social information during Mandarin Chinese accent perception; whereas for a listener with a strong association between Chinese faces and foreign accents we would expect a relatively larger benefit of social information during Mandarin Chinese accent perception.[3]

[3] Exemplar theory is not the only model of speech perception that successfully incorporates the social priming phenomenon. Of particular note, the ideal adaptor model (Kleinschmidt & Jaeger, 2015; Kleinschmidt, Weatherholtz, & Jaeger, 2018) also integrates sociolinguistic inferences such as social priming effects. The ideal adaptor model is a probabilistic learning model that assumes listeners track "lawful variability" (Elman & McClelland, 1983) in the speech signal – leveraging systematic variation (such as socio-indexical cues) during perception to facilitate speech processing. 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 L1- and L2-accented speech.

One of the aims of the present study is to test 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 exemplar theory. To this end, we incorporate individual difference measures of implicit associations alongside our social priming experiments.

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, subjects wishing to maintain a non-prejudiced outward appearance may avoid expressing negative evaluations of L2 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 for an individual 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 (e.g., for examining implicit associations 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 associations 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 by Babel and Russell (2015). As discussed above, Babel and Russell found that Canadian listeners had poorer recognition accuracy for L1-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 associations 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 et al. (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 L2 accent as compared to L1 accent, regardless of the overall intelligibility level of speech, but this difference in audiovisual benefit does not appear to be caused by listeners' implicit biases against Korean talkers.

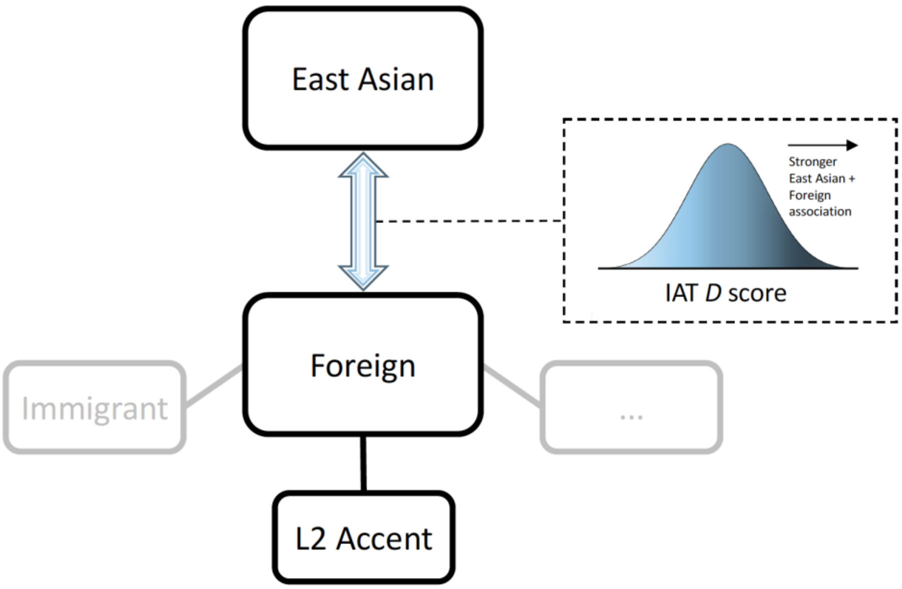
The findings of McLaughlin and colleagues (2022) suggest that the initial correlation found by Yi and colleagues (2013) between IAT scores and audiovisual integration may have been a spurious outcome due a small sample size. Indeed, although IATs are widely used and validated as measures of implicit 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 study, we also use a conservative power analysis to estimate a sample size that would ensure sufficient power to examine relationships between IAT scores and social priming.

Most notable for the present study, although IATs are often framed as measures of individuals' biases, at the most basic level an IAT is a measure of the strength of implicit associations between a set of constructs and social categories. In Yi et al. (2013) the authors used an IAT as a measure of bias and predicted that stronger associations in the predicted direction (i.e., between White faces and the construct American and between Asian faces and the construct Foreign) would correlate with poorer audiovisual integration. In other words, the authors predicted/concluded that negative biases against Asians were interfering with the integration process for Korean-accented English speech (in which video of the Korean speakers was visible). In the present study, although we acknowledge that IATs do index individuals' negative biases, we aim to use the IATs as measures of the strength of implicit associations between constructs. We do not predict that larger IAT scores (stronger biases) will correlate with poorer speech perception for minority race/ethnicity speakers. Rather, we predict that larger IAT scores (stronger associations in the predicted direction) will correlate with larger social priming effects; specifically, a larger difference in performance between a White prime and a minority race/ethnicity prime (see Hypothesis 3, below). Additionally, we include a second IAT (of Good vs. Bad associations) in addition to the American vs. Foreign IAT used by Yi et al. (2013). Our aim by doing so was to determine whether the American vs. Foreign IAT was capturing a unique source of variance in the speech perception data by tapping into listeners' expectations about foreignness and L2 accents as opposed to general positive/negative associations captured by Good/Bad.

Figure 1 visualizes the predicted relationship between IAT scores and social priming. Based on an exemplar framework (Hintzman, 1984; Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001; Johnson, 2006), we propose that non-auditory social factors, such as perceived racial and ethnic characteristics of the speaker, are also stored in exemplars. Over time, abstracted categories are developed that systematically link social information and phonetic patterns. Listeners who have developed stronger implicit associations between a given racial/ethnic category and the construct foreign (which indexes L2 Accent) ought to also be more greatly affected by this racial/ethnic information during speech processing. In cases in which the L2 accent receiving greater activation matches the incoming speech signal, this would facilitate speech processing; when the incoming signal and accent (e.g., an L1 accent) do not match, however, this would hinder speech processing.

Figure 1

Diagram of Predicted Relationship Between IAT D Scores and Social Priming Effects



Note. An example of abstracted nodes systematically linking social information and phonetic patterns is shown for the category East Asian. The strength of associations between the nodes for East Asian and Foreign is visualized with a double-sided arrow of variable thicknesses (where a thicker arrow indicates a stronger, more facilitative, connection). We hypothesize that the strength of this connection can be measured with implicit association test (IAT) *D* scores; a stronger connection (thicker arrow) is doubly represented with varying shades of blue corresponding to an example IAT *D* score distribution. In this example, larger *D* scores (darker blue) indicate stronger associations between the category East Asian and the construct Foreign. Thus, an individual with a larger IAT *D* score is predicted to have a stronger, more facilitative, connection between the East Asian and Foreign nodes. Critically, the Foreign node is assumed to have further connections that *include L2 accent* (based on prior stereotyping research, Zou & Cheryan, 2017), which would be facilitated, second-hand, whenever the Foreign node is facilitated. Based on this model, listeners who have developed stronger implicit associations between the East Asian category and the construct Foreign (which indexes L2 Accent) ought to also be more greatly affected by this racial/ethnic information during speech processing. Note that although East Asian is shown in this example, the same relationship is predicted for Latinx.

One perplexing assumption that follows from the prediction illustrated in **Figure 1** is that listeners may benefit from racial/ethnic information about the speaker even when they do not know the specific L2 accent they are listening to. Indeed, if a listener perceives a speaker to be East Asian but does not know their nationality, how are they to know whether to prepare for a Mandarin accent versus a Japanese accent (and so on)? One possibility originally proposed by Melguy and Johnson (2021) is that listeners are not directionally shifting their phonemic categories in anticipation of a specific L2 accent, but rather relaxing their categories (see similar discussion by Zheng & Samuel, 2020). On this view, **Figure 1** may be interpreted as listeners preparing for any kind of L2 accent upon activation of the L2 accent node. The bounds of these types of accent-general adaptations (i.e., can an East Asian face prepare a listener for Spanish accent?) remains an open empirical question (see McLaughlin & Van Engen, 2023b).

Language and social attitudes

Not all accents are viewed equally among language users. Speakers with non-standard accents – and particularly non-standard accents of lower status – often face prejudice and stigmatization that impacts their everyday lives (Carlson & McHenry, 2006; Purnell, Idsardi, & Baugh, 1999). For L2-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 L2 and regional speakers report experiences of stigmatization, in the United States speakers with L2 accents report more communication difficulties and a lesser sense of belonging (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 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' reported attitudes are typically influenced by speech comprehensibility. For example, more comprehensible L2-accented speakers tend to be rated more positively on status traits than less comprehensible L2-accented speakers (Bresnahan et al., 2002). What's more, Dragojevic and Giles (2016) found for L1 English listeners that even when the same L2 speaker was presented, but either in quiet or in white noise (the latter of which was less comprehensible), subjects rated the L2 accent in noise lower for status traits and reported poorer affect (i.e., poorer mood).

Such stigmatization of non-standard accents has real world impacts. Listeners 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. While self-reported attitudes provide valuable insight into the stigmatization of accented speech, they 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.

The current study

Across three experiments, we examine the effects of social information on the perception of speech presented in noise and how individual differences in implicit associations may predict these effects. In our first two experiments, we focus on social priming in the context of L1 accent perception; in other words, we predict *negative* social priming effects (similar to Babel & Russell, 2015). Specifically, we examine a White vs. an East Asian prime (Experiment 1) and a White vs. a Latina prime (Experiment 2). In our second experiment, we specifically decided to include a Latina prime because Latinx primes have not been examined in an experiment of this type previously. To our knowledge, Vaughn (2019) is the only study to have examined the effects of a Latinx guise (presented orthographically) on perception of speech, and the focus of that study was L2 accent perception (whereas the present study focuses on L1 accent perception). Vaughn found that adaptation to L2 Spanish-accented English was better when listeners were given a guise implying that the speaker was Latino, as opposed to no information at all.[4] To the best of our knowledge, this is the first study to examine whether a Latina prime may elicit a similar negative social priming effect as an East Asian prime.

[4] Surprisingly, the results of Vaughn (2019) indicated that listeners given a guise suggesting the speaker was Latino but had an L1 accent actually did better than listeners given a guise suggesting the speaker was Latino and had an L2 accent (the latter of which was the accurate description). This outcome indicates the even less specific guises can assist the listener during perception.

In a third experiment, we examine social priming in the context of an L2 accent: Mandarin Chinese-accented English. This experiment parallels the design of McGowan (2015), but compares the effects of a White versus and East Asian prime on L2 accent perception within subjects (as opposed to between subjects).

In all experiments, participants completed two blocks of a speech transcription task with a different image of the (supposed) speaker presented on screen in each block. Participants also completed two IATs (American vs. Foreign associations and Good vs. Bad associations) and measures of their affect and attitudes toward the speakers. Our inclusion of the Good-Bad IAT in addition to the American-Foreign IAT was largely exploratory. We predicted (as listed in Hypothesis 4, below) that the American-Foreign IAT may capture a unique source of variance in the social priming data because of a relationship specifically between the construct of foreignness and expectations about L2 accent. If this was the case, including the Good-Bad IAT would provide useful discriminant validity.

Here, we outline the overarching and experiment-specific hypotheses of the present study:

1) For Experiments 1 and 2: Race/ethnicity primes (images of a White, East Asian, vs. Latina face) will affect the perception of American, L1-accented English in babble. Specifically, subjects will have better transcription accuracy for the White prime than the East Asian or the Latina prime.
2) For Experiment 3: Race primes (images of a White vs. an East Asian face) will affect the perception of Mandarin Chinese, L2-accented English in babble. Specifically, subjects will have better transcription accuracy for the East Asian prime than the White prime (as in McGowan, 2015; cf., Kutlu et al., 2022).

3) For all experiments: Individual listeners' implicit associations will be related to the size of their "social priming effects" (Figure 1). The size of the difference between conditions (e.g., transcription performance for a White versus an East Asian prime) will vary by listener, and we expect that listeners with larger priming effects will have stronger implicit associations between these races/ethnicities and the constructs American vs. Foreign.
4) For all experiments: The relationship between social priming costs and implicit associations will depend on the type of IAT. Specifically, we expect that the Good-Bad IAT will not significantly predict social priming costs, while the American-Foreign IAT will, confirming a unique relationship between the construct of foreignness and social priming effects.

Experiment 1

In Experiment 1, we examine the effects of a White versus an East Asian prime on the perception of American-accented English speech for White American subjects. Additionally, we explore two types of implicit racial biases using IATs, as well as explicit attitudes, and investigate whether these bias measures predict social priming for L1 accent.

Method

Transparency and openness

This study complies with transparency and openness guidelines. Pre-registration of both Experiments 1 and 2 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. Data was collected in 2022, which may constrain the generalizability of the findings given that social attitudes and speech patterns change over time.

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, 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 $r = .15$.

Participants

Notably, in our recent work (McLaughlin & Van Engen, 2023b), data from a representative, convenience sample of American college subjects was collected, and we explored the effect of listeners' race on social priming effects. The results of the analysis indicated social priming effects in the sub-sample of White subjects, but not in the sub-sample of non-White subjects. Based on these findings, in the present study we opted to focus on social priming effects in White American listeners.

Young adult subjects (age mean = 24.5; age range = 18-35; gender: 268 female, 114 male, seven non-binary) were recruited using the website Prolific to participate in Experiment 1 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 subjects would need to be excluded from the sample, and thus planned to recruit up to a maximum of 400 subjects (i.e., the maximum our budget allowed), or until 350 valid subjects participated. In total, 389 subjects participated in the experiment, 38 of whom were excluded for one or more of the following reasons: failing to meet eligibility criteria (despite Prolific's pre-screening; three), failing the headphone screening (up to two attempts allowed; 16), self-reporting using speakers instead of headphones for any task (three), failing attention-check trials in the speech transcription task (one), performing greater than or equal to three standard deviations away from the group average in the speech transcription task (eight), or self-reporting that their data should be excluded (seven). All subjects correctly identified the race/ethnicity of priming images in the experiment (details below), and no subjects needed to be excluded for this reason.[5] The final *N* of the experiment was slightly above the target: *N* = 351.

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 White prime image, 96.2% of responses were coded as White or Caucasian and the remaining

[5] Given the open-response nature of the question, some subjects provided responses that were not viable (e.g., "American" for a White prime). We determined not to exclude subjects with non-viable responses on the assumption that the question may have been misunderstood.

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.

Materials

Speech perception task

Auditory stimuli were retrieved from SpeechBox (Bradlow, n.d.), and included recordings of two female American-accented L1 speakers of English reading aloud 40 Hearing-In-Noise-Test sentences (speakers KEI_EF04 and KEI_EF05; Nilsson, Soli, & Sullivan, 1994; Bradlow, Blasingame, & Lee, 2018). 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 L1 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, two additional audio files were recorded by a different L1 speaker of American English. 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, pictures of a White female and an East Asian female were selected from the Chicago Face Database (CFD; files CFD-WF-003-003-N.jpg and CFD-AF-253-130-N.jpg, respectively; Ma, Correll, & Wittenbrink, 2015). The CFD’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 two women featured in the images selected for Experiment 1 were similarly rated for attractiveness, neutrality of expression, and high prototypicality of race in the CFD’s metadata.

Implicit association tests (IATs)

For both IATs, images of White and East Asian 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 and were approximately matched (on average) for attractiveness, neutrality of expression, and high prototypicality of race. 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. All images were presented in color.

The American-Foreign IAT included images of eight American scenes and eight foreign scenes, representing the constructs *American* and *Foreign*, respectively (all photos from Yi et al., 2013). American scenes included: the White House, the Golden Gate Bridge, the Pentagon, the Statue of Liberty, the Liberty Bell, the US Capitol Building, the Empire State Building, and Central Park (New York). Foreign scenes included: the Taj Mahal, Stonehenge, the Great Wall of China, the leaning Tower of Pisa, the Sydney Opera House, the Great Pyramids, the Eiffel Tower, and Angkor Wat. 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.

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 1. The screening leverages the perceptual phenomenon Huggins Pitch, which can only be detected when stimuli are presented dichotically (and, thus, cannot be perceived over loudspeakers). In the task, subjects are presented three noise bursts each trial, one of which contains a hidden tone. 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. 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 the correct response (“the boy fell from the window”), and were instructed 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 throughout the transcription task. We decided to present the priming images in a block-wise fashion as was done in McGowan (2015) after early piloting attempts with randomized trial-to-trial priming were unsuccessful (see McLaughlin & Van Engen, 2023a).

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 block. The presentation order of primes and the combination of each prime and each of the two female speakers was counterbalanced across subjects (four counterbalanced versions 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 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 1

Summary of the Implicit Association Test (IAT)

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 (i.e., the main blocks containing data for analysis). 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 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, 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 estimate what percentage of their social network included East Asian or Latinx individuals, respectively; 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 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). The settings for the autoscore function were set to allow common misspellings (the Acceptable Spell Rule), omittance or unnecessary inclusion of double letters (the Double Letter Rule), differences in tense (the Tense and Tense+ Rules), and differences in plurality (the Plural Rule). Word order did not affect scoring. 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).

One issue that we predicted could occur is an order effect on social priming. That is, after subjects become aware that we are asking questions about the speakers, and that the two speakers differ in race or ethnicity, they may suspect that their racial/ethnic biases are being examined. Thus, we also pre-registered an analysis of the effect of counterbalancing, and decided to conduct separate between-subject analyses of just the first block of the task, where necessary.

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 adapted to the new sorting rule).

Results

Direct analyses of the IATs are reported in **Supplemental Materials B**, and direct analyses of the affect and attitudes questionnaire are reported in **Supplemental Materials C**. The relationship between subjects' IAT *D* scores from the American-Foreign and Good-Bad IATs and their performance in the speech transcription task are examined as part of the analyses of the speech transcription task below.

Pre-registered analyses

Reliability of speech transcription tasks

Cronbach's alpha of the speech transcription task was 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 (α = .84).

Variance inflation factors

The variance inflation factors of the main effects in the model were checked to confirm that there were no multi-collinearity issues. 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 2** summarizes the variance inflation factors for all experiments. For Experiment 1, 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 (Craney & Surles, 2002).

Table 2

Variance Inflation Factors (VIFs) for All Experiments

Fixed Effect	Experiment 1 VIFs	Experiment 2 VIFs	Experiment 3 VIFs
Prime	1.00	1.00	1.00
AF *D* score	1.06	1.10	1.13
GB *D* score	1.06	1.10	1.13
Counterbalance	1.00	1.00	1.01

Primary analyses

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 (a full target sentence), and the predicted variable of the models was two columns: count of correctly identified target words and count of missed target words.[6] Although there are multiple keywords per sentence/trial, the GLMER model is nonetheless able to predict both the counts of the correct versus missed target words using a binomial regression. 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, and counterbalance (dummy-coded levels: EA-W, W-EA), as well as interactions between prime and each additional predictor.[7] Counterbalance EA-W contained subjects who were presented with the East Asian prime block first, and Counterbalance W-EA 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**). A summary of the models and their R syntax is provided in **Appendix A** for Experiment 1. Most notably, the effect of prime did not significantly improve model fit ($\chi^2 < 0.01$, $DF = 1$, $p = .96$), contrary to our predictions (**Figure 2A**). Performance for each priming condition in all experiments is summarized in **Table 4**.

---

[6] A simplified example of the R code used to predict two columns simultaneously with GLMER would be:

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime (1 + Prime | Subject) + (1 + Prime | Item),
data = E1_data, family = "binomial")
```

[7] Our pre-registered analysis originally planned to include the measures of affect and attitudes in the full model. However, based on feedback during peer-review we decided to simplify the models (presenting those outcomes in the Supplemental Materials, instead). Results remained the same.

Table 3

Log-likelihood Model Comparisons for Experiment 1

Effect	$\chi^2(1)$	*p*
Prime	< 0.01	.96
American-Foreign *D* score	4.23	.04
Good-Bad *D* score	0.32	.57
Counterbalance	1.96	.16
Prime x American-Foreign *D* score	0.64	.43
Prime x Good-Bad *D* score	0.36	.55
Prime x Counterbalance	53.49	< .001

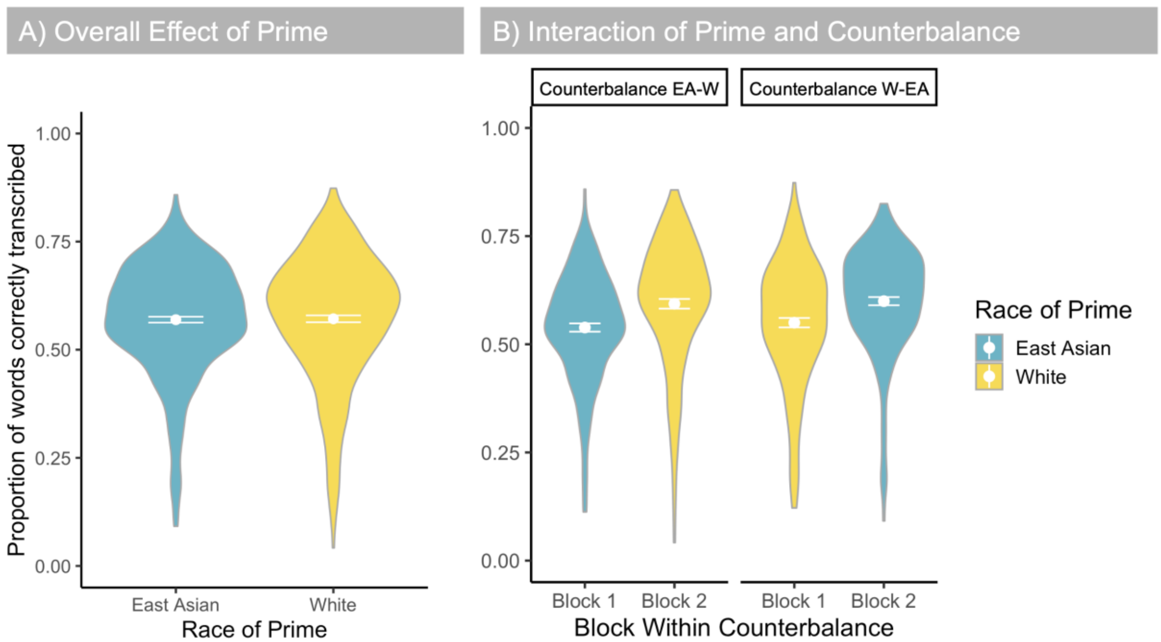
Table 4

Average Performance for All Experiments

Prime	Experiment 1	Experiment 2	Experiment 3
White	57.2%	57.2%	58.1%
East Asian or Latina	56.9%	57.3%	60.3%

Figure 2

Effects of Prime and Counterbalance in Experiment 1

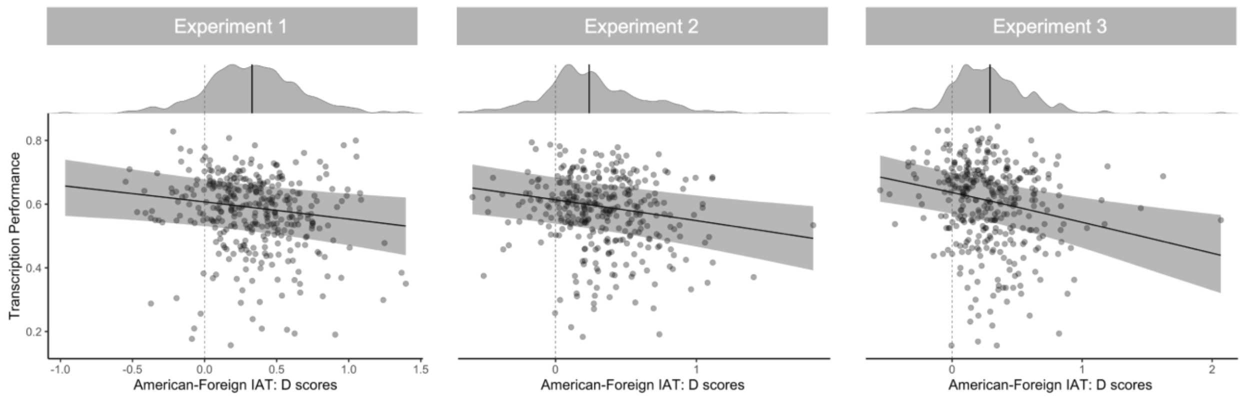


Note. A) Violin plots, means, and standard error bars show the non-significant effect of prime (i.e., viewing an East Asian versus White face) on transcription accuracy for American-accented English speech. The y-axis is a summary measure of performance, showing the proportion of keywords transcribed accurately per sentence (averaged across trials). B) Violin plots, 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 EA-W were presented with the East Asian prime block first, while subjects in Counterbalance W-EA were presented with White prime block first.

Individual differences in American-Foreign *D* scores improved model fit (χ^2 = 4.23, *DF* = 1, *p* = .04) but individual differences in Good-Bad *D* did not (χ^2 = 0.32, *DF* = 1, *p* = .57). For the former, the model estimate indicated that subjects who performed worse on the task had significantly larger *D* scores (i.e., stronger associations between White and American and between East Asian and Foreign) for the American-Foreign IAT (*ß* = -0.24; **Figure 3**).

Figure 3

Relationship Between Performance and American-Foreign D scores Across Experiments



Notes. Visualization of the relationship between performance on the speech perception task (summarized as proportion of keywords correctly transcribed) and individual differences in American-Foreign IAT *D* scores for each experiment. In lower plots, subjects are represented as individual points with a model-predicted fit line overlaid with 95% confidence interval, and, in upper plots, 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 counterbalance was not significant ($\chi^2 = 1.96$, $DF = 1$, $p = .16$), but the interaction between prime and counterbalance did improve model fit ($\chi^2 = 53.49$, $DF = 1$, $p < .001$). As visualized in **Figure 2B**, this interaction is driven by improvement from Block 1 to Block 2, resulting in better performance on the White prime for Counterbalance EA-W and better performance on the East Asian prime for Counterbalance W-EA. The interactions between prime and each of the IAT measures were non-significant (both p's > .05).

Follow-up analyses: Block 1 data

We anticipated that the order of presentation of the race primes (counterbalancing) may affect the social priming effect because of listeners' awareness of the manipulation of race (i.e., the image onscreen changes either to or from a minority race, possibly signaling to the participant that the experiment is examining racial biases). To examine this possibility, we pre-registered an additional set of analyses. Following the same analysis plan, we constructed a set of models comparing data only from the first block of the speech transcription task. In this way, we were able to isolate the analyses to a portion of the experiment in which subjects were not aware of the social priming manipulation. 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, and interactions between prime and these two other effects. Log-likelihood model comparisons are summarized in **Table 5**.

Table 5

Log-likelihood Model Comparisons for Block 1 of Experiment 1

Effect	$\chi^2(1)$	*p*
Prime	0.01	.92
American-Foreign *D* score	4.87	.03
Good-Bad *D* score	0.68	0.41
Prime x American-Foreign *D* score	0.22	0.64
Prime x Good-Bad *D* score	6.41[a]	0.01[a]

Note. Values marked with superscript *a* are non-significant after removing subjects with outlying (three standard deviation from the mean) Good-Bad *D* scores.

As was the case for the within-subject analysis, the effect of prime did not significantly improve model fit (χ^2 = 0.01, *DF* = 1, *p* = .92). Thus, it does not appear that the effect of block was reducing our ability to detect an effect of social priming. Results of this analysis remained largely the same as those in the primary analysis with the exception of a significant interaction between prime and Good-Bad *D* scores (χ^2 = 6.41, *DF* = 1, *p* = .01). 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 significant (*p* = .29). Thus, we did not examine this interaction further.

Exploratory analyses

Bayesian analyses

To further investigate the null outcome of the social priming manipulation, we conducted a Bayesian pair-wise t-test in JASP (JASP Team, 2023). The predicted variable in this analysis was the mean proportion of keywords correctly identified by subject for each of the prime conditions (White and East Asian). Settings selected for the Student's t-test with an alternative hypothesis of White ≠ East Asian. The default prior was used (Cauchy prior width = 0.707) and a Bayes factor robustness check was run to compare the selected prior against a wide and ultrawide prior. Results of the t-test with the default prior indicated strong evidence in favor of the null hypothesis ($BF_{10} = 0.06$); the wide ($BF_{10} = 0.04$) and ultrawide ($BF_{10} = 0.03$) priors from the robustness check indicated progressively stronger evidence in favor of the null hypothesis. Descriptions of the Bayes factors (e.g., "strong evidence") are based on the JASP guidelines (van Doorn et al., 2021).

Interim discussion

In Experiment 1, we did not find 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 following an East Asian prime than it was following a White prime, and additional Bayesian statistics confirmed that there was strong evidence in favor of the null hypothesis. We also found little evidence to indicate a 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. Only *D* scores from the American-Foreign IAT significantly predicted overall performance on the task, indicating that listeners with stronger White + American and East Asian + Foreign associations had poorer transcription accuracy. However, the interaction between these IAT *D* scores and priming was non-significant.

Altogether, the results of Experiment 1 do not indicate that race information affects listeners' recognition of American-accented English speech. Although listeners with stronger implicit racial biases performed worse during the speech perception tasks, this relationship was unrelated to the race of the speaker presented on the screen during the task. This disconnect in outcomes may indicate that the implicit association tests are predicting overall listening performance due to a third variable (discussed further in the General Discussion). In Experiment 2 (collected in tandem with Experiment 1), we sought to examine these same topics for another set of social primes.

Experiment 2

In Experiment 2, we examine the effects of a White versus a Latina prime on perception of American-accented English for White American subjects and explore the relationship between social priming and implicit and explicit biases. By including a Latina prime in Experiment 2, we aimed to examine negative social priming effects for an ethnicity that had not been examined in prior research. Data for Experiment 2 was collected in tandem with data for Experiment 1[8], before the results were known.

Method

Participants

Young adult subjects (age mean = 24.5; age range = 18-35; gender: 283 female, 92 male, 14 non-binary) were recruited using the website Prolific to participate in Experiment 2 online. Participants who participated in Experiment 1 were not able to enroll in Experiment 2. Inclusion criteria matched Experiment 1: Prolific's demographic filters selected for White young adults

[8] Although it would certainly be possible to examine Experiment 1 and 2 jointly (i.e., with the different priming manipulations presented as a condition manipulation), our a-priori plan was to treat them as separate experiments. This design choice was motivated by stimuli differences in the IATs used for the subjects in Experiment 1 versus the subjects in Experiment 2.

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 once again anticipated that some subjects would need to be excluded, and thus planned to recruit up to 400 subjects (i.e., the maximum our budget allowed), or until 350 valid subjects participated. In total, 389 subjects participated in the experiment, 36 of whom were excluded for one or more of the following reasons: failing to meet eligibility criteria (despite Prolific's pre-screening; three), failing the headphone screening (up to two attempts allowed; 17), self-reporting using speakers instead of headphones for any task (three), failing attention-check trials in the speech transcription task (one), performing greater than or equal to three standard deviations away from the group average in the speech transcription task (seven), or self-reporting that their data should be excluded (five). In addition to these pre-registered exclusions, we decided to remove any subjects who misidentified the race/ethnicity of the prime images (see section below for full details). Only subjects who provided an incorrect response (e.g., "Asian" for the White prime image) were removed from analyses; subjects who made responses that were not clear (e.g., "American" for the White prime image) were retained. This resulted in an additional exclusion of 15 participants. The final N of the experiment was slightly below the target: $N = 338$.

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"). Before removing the 15 participants noted above, the accuracy of responses were as follows: 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.

Procedure

Procedures for Experiment 2 matched Experiment 1, with the only exception being the images used in the speech perception task (CFD file CFD-LF-255-088-N.jpg was used for the Latina prime) and the IATs. For the speech perception task in Experiment 2, the image of the White female that was selected for Experiment 1 was used again, along with a novel image of a Latina female selected from the CFD (Ma et al., 2015). The two women were similarly rated for attractiveness, neutrality of expression, and high prototypicality of race in the CFD’s metadata. For the IATs, the same eight White faces (four male, four female) from Experiment 1 were used again, along with four male and four female Latinx faces that were approximately matched for attractiveness, neutrality of expression, and high prototypicality of race. The faces used in the IATs did not overlap with those selected for the speech perception task, and all images were presented in color.

Results

Direct analyses of the IATs are reported in **Supplemental Materials B**, and direct analyses of the affect and attitudes questionnaire are reported in **Supplemental Materials C**. The relationship between subjects’ IAT *D* scores from the American-Foreign and Good-Bad

IATs and their performance in the speech transcription task are examined as part of the analyses of the speech transcription task below.

Pre-registered analyses

Reliability of speech transcription task

As in Experiment 1, Cronbach's alpha of the speech transcription task was 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 (α = .78).

Variance inflation factors

As reported in **Table 2**, we checked the variance inflation factors (i.e., multi-collinearity) of the main effects in the within-subject model for Experiment 2. 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 Experiment 1, in Experiment 2 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 (Craney & Surles, 2002).

Primary analyses

Model specifications matched those in the analysis of Experiment 1 with the exception of the random effects. Model comparisons indicated that random slopes for speaker did not improve model fit, and were thus dropped from the random effect structure. Fixed effects in the *glmer()* model included: prime (dummy-coded levels: Latinx, White), American-Foreign *D* score, Good-Bad *D* score, and counterbalance (dummy-coded levels: Counterbalance L-W, Counterbalance W-L), as well as interactions between prime and each additional predictor. Counterbalance L-W contained subjects who were presented with the Latina prime block first, and Counterbalance W-

L 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 6**). A summary of the models from Experiment 2 is provided in **Appendix B**. As in Experiment 1, the effect of prime did not significantly improve model fit ($\chi^2 < 0.01$, *DF* = 1, $p = .96$; **Figure 4A**).

Table 6

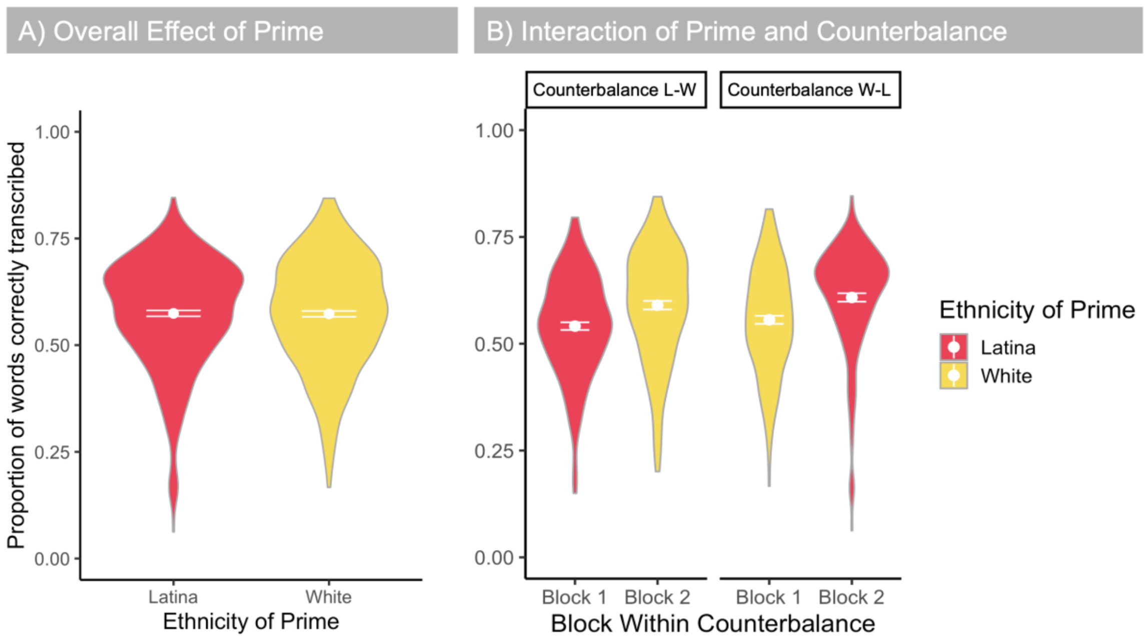
Log-likelihood Model Comparisons for Experiment 2

Effect	$\chi^2(1)$	*p*
Prime	< 0.01	.96
American-Foreign *D* score	4.01	< .05
Good-Bad *D* score	7.64	.006
Counterbalance	2.37	.12
Prime x American-Foreign *D* score	1.68	.20
Prime x Good-Bad *D* score	4.02[a]	< .05[a]
Prime x Counterbalance	49.80	< .001

Note. Values marked with superscript *a* are non-significant after removing subjects with outlying (three standard deviation from the mean) Good-Bad *D* scores.

Figure 4

Effects of Prime and Counterbalance in Experiment 2

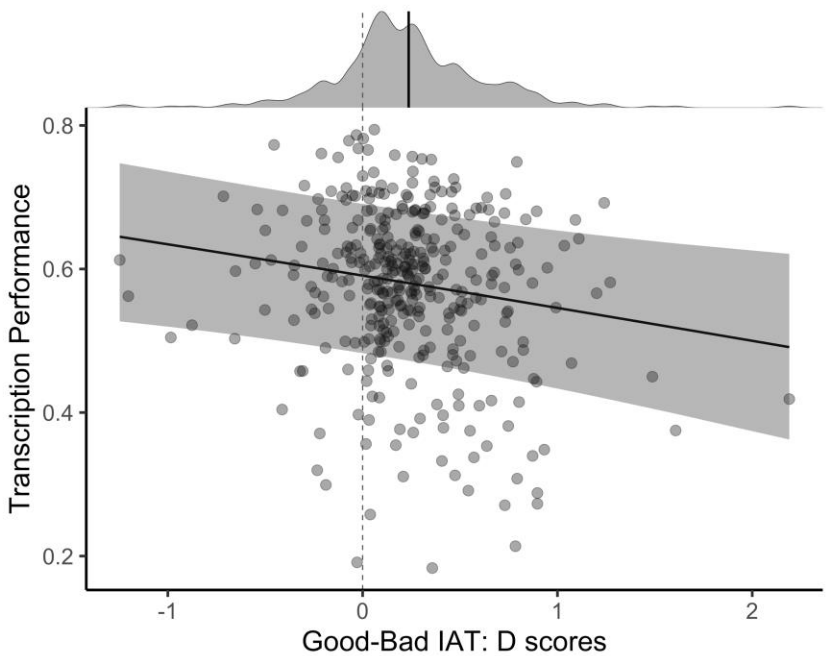


Note. A) Violin plots, means, and standard error bars show the non-significant effect of prime (i.e., viewing a Latina versus White face) on transcription accuracy. The y-axis shows the proportion of keywords transcribed accurately per sentence (averaged across trials). B) Violin plots, means, and standard error bars 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 L-W were presented with the Latina prime block first, while subjects in Counterbalance W-L were presented with White prime block first.

For the IATs, individual differences in American-Foreign *D* scores (χ^2 = 4.01, *DF* = 1, *p* < .05; **Figure 3**) and Good-Bad *D* scores (χ^2 = 7.64, *DF* = 1, *p* = .006) both significantly improved model fit. The model estimates (*ß* = -0.20 and *ß* = -0.22, respectively) indicated that subjects with larger *D* scores performed more poorly overall on the speech transcription task. **Figure 5** 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 5

Relationship Between Performance and Good-Bad D scores in Experiment 2



Note. 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 Experiment 2. In lower plot, subjects are represented as individual points with a model-predicted fit line overlaid with 95% confidence interval, 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.

As in Experiment 1, the effect of counterbalance did not improve to model fit (χ^2 = 2.37, *DF* = 1, *p* = .12), but the interaction between prime and counterbalance did make a significant contribution (χ^2 = 49.80, *DF* = 1, $p < .001$). The model estimate for the interaction indicated a large improvement from Block 1 to Block 2 (*ß* = -0.49; **Figure 4B**). Lastly, we examined the interaction between each set of *D* scores and the effect of prime. The interaction between American-Foreign *D* scores and prime was not significant (χ^2 = 1.68, *DF* = 1, p = .20), but the interaction between the Good-Bad *D* scores and prime was (χ^2 = 4.02, *DF* = 1, $p < .05$). However, follow-up analyses indicated that this trend was driven by six participants with outlying Good-Bad *D* scores (defined as greater than three standard deviations from the mean). When these outliers were removed from the dataset this interaction was no longer significant (χ^2 = 2.68, *DF* = 1, p = .10).

Follow-up analyses: Block 1 data

Given the significant interaction between prime and counterbalance, we conducted a follow-up analysis using data from the first block of the speech transcription task only. For this between-subject analysis, our sample included 170 subjects in the Latina prime group and 168 subjects in the White prime group. Random and fixed effects matched those in the between-subject analysis in Experiment 1. Log-likelihood model comparisons are summarized in **Table 7**.

Table 7

Log-likelihood Model Comparisons for Block 1 of Experiment 2

Effect	$\chi^2(1)$	*p*
Prime	0.01	0.91
American-Foreign *D* score	3.57	.06
Good-Bad *D* score	8.73	.003
Prime x American-Foreign *D* score	1.86	.17
Prime x Good-Bad *D* score	0.07	.80

The results remained largely the same as for the within-subject analysis, with the exception of the interaction between prime and Good-Bad *D* scores, which was non-significant even when outliers were retained (χ^2 = 0.07, *DF* = 1, *p* = .80). Note also that the effect of American-Foreign *D* scores was only marginal (*p* = .06) in this follow-up analysis. Most importantly, this analysis of the Block 1 data indicated no effect of prime (χ^2 = 0.01, *DF* = 1, *p* = .91). Thus, it does not appear that subjects' potential awareness of the priming manipulation can account for the lack of social priming effect.

Exploratory analyses

Bayesian analysis

As in Experiment 1, we conducted an exploratory analysis of the null outcome for the social priming manipulation using a Bayesian pair-wise t-test in JASP (JASP Team, 2023). The predicted variable was the mean proportion of keywords correctly identified by subject for each of the prime conditions (White and Latinx). Settings selected for the Student's t-test with an alternative hypothesis of White ≠ Latinx. The default prior was used (Cauchy prior width =

0.707) and a Bayes factor robustness check was run to compare the selected prior against a wide and ultrawide prior. Results of the t-test with the default prior for Experiment 2 indicated strong evidence in favor of the null hypothesis (BF_{10} = 0.06); the wide (BF_{10} = 0.04) and ultrawide (BF_{10} = 0.03) priors from the robustness check indicated progressively stronger evidence in favor of the null hypothesis.

Combined analysis of speech transcription tasks across experiments

Based on the analyses of the Block 1, it appeared that the insignificant effects of prime were, nonetheless, consistent in direction and size across experiments. Thus, we decided to examine the effect of prime within Block 1 in a combined dataset using generalized linear mixed-effects regression, 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 = 689; 345 White, 344 Minority). Random effects included intercepts by subject and by item. Only the fixed effect of prime was entered into the model, and it did not improve model fit (χ^2 = 0.02, DF = 1, p = .90), reaffirming the null outcomes of the primary analyses.

Interim discussion

Our examination of social priming in Experiment 2 did not reveal an effect of perceived ethnicity on the perception of American-accented English speech presented in noise. Specifically, transcription accuracy was not significantly different following a Latina prime than it was following a White prime. This was further affirmed by Bayesian statistics, which indicated strong evidence in favor of the null hypothesis.

Individual differences in implicit racial associations (i.e., associations with the constructs American vs. Foreign and the constructs Good vs. Bad, respectively) were unrelated to the effect of social priming in the speech perception task, matching the outcomes of Experiment 1.

However, for both of these measures significant relationships emerged with overall performance on the task: Listeners with stronger White + American/Good and East Asian + Foreign/Bad associations had poorer overall transcription accuracy.

Experiment 3

In Experiment 3, we conducted a follow-up investigation of social priming effects for *L2-accented* speech. Our aim was to determine whether a relationship between implicit racial associations and social priming would emerge in the context of L2 accent. Additionally, we suspected that the null priming effects observed in Experiments 1 and 2 may have prevented a significant interaction with the IAT measures; in other words, by examining an accent for which a social priming effect will (presumably) be present, we should be better situated to observe an interaction with implicit racial associations. We decided to mirror the design of Experiment 1, and predicted that for perception of L2 Mandarin accent, subjects would perform better when viewing an East Asian prime than a White prime, and that implicit racial associations would predict the size of a given subject's social priming effect. We report the findings of Experiment 3 below, which matched Experiment 1 in all aspects except for the accent of the target talker in the audio files.

Method

All stimuli and procedures for Experiment 3 matched Experiment 1 with the exception of the auditory stimuli. We report details of these stimuli below, as well as information about the sample of participants and exclusions.

Participants

Adult subjects were recruited using the website Prolific to participate online. Subjects who completed Experiments 1 or 2 were not able to enroll in the study. Due to a mistake in the initial posting of the experiment on Prolific, a sample of subjects any age over 18 was initially collected (instead of only young adult subjects ages 18-35). We analyze and report results for middle-aged and older adults in the **Supplemental Materials**, and focus on the young adult analysis for Experiment 3. Prolific's demographic filters were set to include White adults 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). A total of 586 subjects were recruited for Experiment 3 (age mean = 34.31; age range = 18-75). After separating the middle-aged and older adults into the supplemental dataset (*n* = 165) and excluding subjects (*n* = 71), 350 young adult subjects remained for the Experiment 3 analyses (age mean = 26.85; age range = 18-35; gender: 221 female, 116 male, 13 non-binary). The 71 subjects excluded from the sample were removed for one or more of the following reasons: failing to meet eligibility criteria (despite Prolific's pre-screening; eight), failing the headphone screening (up to two attempts allowed; 35), self-reporting using speakers instead of headphones for any task (six), failing attention-check trials in the speech transcription task (nine), performing greater than or equal to three standard deviations away from the group average in the speech transcription task (three), or self-reporting that their data should be excluded (10). In addition to these pre-registered exclusions, we decided to remove any subjects who misidentified the race/ethnicity of the prime images (see section below for full details). Only subjects who provided an incorrect response (e.g., "Asian" for the White prime image) were removed from analyses. Subjects who made responses that were not clear (e.g., "American" for the White prime image) were retained. This resulted in an additional exclusion of two young adult participants. The final *N* of the experiment was: *N* = 348.

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”). Before removing the two participants noted above, the accuracy of responses were as follows: the White prime was coded as White or Caucasian for 83.9% of responses, Asian for 0.2% of responses, Latinx or Hispanic for 0.2% of responses, and RNS for 15.8% of responses; the East Asian prime was coded as Asian for 98.2% of responses, White or Caucasian for 0.4% of responses, and RNS for 1.5% of responses. Note that the amount of RNS responses for the White prime in Experiment 3 (15.8%) was greater than in Experiments 1 (3.8%) or 2 (0.8%), which is likely an influence of the combination of the White prime with a non-typical accent (Mandarin Chinese-accented English). A large number of the RNS responses were things such as: Eastern European, German, Russian, Dutch, and so on. When excluding subjects with RNS responses, all results remained the same.

Materials

Auditory stimuli included recordings of two Mandarin Chinese-accented (L1 Mandarin Chinese, L2 English) female speakers of English reading aloud semantically normal sentences developed by Van Engen, Chandrasekaran, and Smiljanic (2012). All sentences were six words in length, with four keywords: “the *gray mouse ate* the *cheese*.” The two Mandarin-accented speakers were selected for the present experiment based on pilot data that indicated they were similarly intelligible in quiet listening conditions (93% and 94% intelligible, respectively). The same samples of background noise used in Experiments 1 and 2 were used for Experiment 3 (i.e., six-talker babble created from 30 simple, meaningful sentences produced by three male and three female L1 speakers of American English; Bradlow & Alexander, 2007). The sentence targets and background babble were mixed at a signal-to-noise ratio (SNR) of -2 dB, with target onset lagging 500 ms after the start of the babble. Piloting of the two speakers indicated that at -2 dB SNR they were both approximately 60% intelligible, roughly matching the difficulty of the speakers presented in Experiments 1 and 2.

For the attention-check trials, the same two audio files used in Experiments 1 and 2 were used. These files were recordings of the sentences “please type a single G” and “please type a single Q” recorded by an L1 speaker of American English (presented without background noise).

Transparency and openness

This study complies with transparency and openness guidelines. Experiment 3 was pre-registered with Open Science Framework: https://osf.io/36v9x. 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.

Results

Pre-registered analyses

Variance inflation factors

As in Experiments 1 and 2, we calculated the variance inflation factors (i.e., multi-collinearity) of the main effects in the within-subject model for Experiment 3 (**Table 2**). None of the factors indicated issues of multi-collinearity (Craney & Surles, 2002).

Primary analyses

Model specifications matched those in the analyses of Experiment 1. Log-likelihood model comparisons are summarized in **Table 8**. A summary of the models from Experiment 3 is provided in **Appendix C**. Most notably, results of the model comparisons indicated that the effect of prime significantly improved model fit (χ^2 = 4.16, *DF* = 1, *p* = .04; **Figure 6A**). As predicted, subjects had better listening performance for the Mandarin Chinese-accented English speech when viewing the East Asian prime than the White prime (*ß* = -0.09).

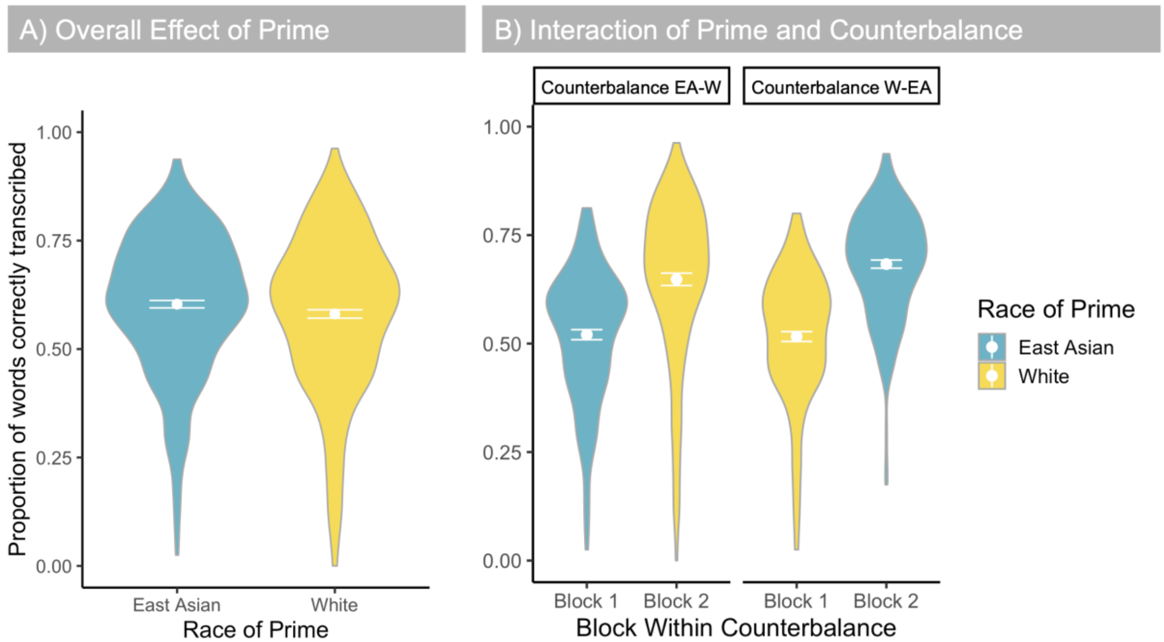
Table 8

Log-likelihood Model Comparisons for Experiment 3

Effect	$\chi^2(1)$	*p*
Prime	4.16	.04
American-Foreign *D* score	4.19	.04
Good-Bad *D* score	2.90	.09
Counterbalance	0.94	.33
Prime x American-Foreign *D* score	2.84	.09
Prime x Good-Bad *D* score	0.13	0.72
Prime x Counterbalance	9.39	.002

Figure 6

Effects of Prime and Counterbalance in Experiment 3



Note. A) Violin plots, means, and standard error bars show the marginal effect of prime (i.e., better performance when viewing an East Asian versus White face) on transcription accuracy for Mandarin-accented English speech. The y-axis is a summary measure of performance, showing the proportion of keywords transcribed accurately per sentence (averaged across trials). B) Violin plots, 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 EA-W were presented with the East Asian prime block first, while subjects in Counterbalance W-EA were presented with White prime block first.

Individual differences in American-Foreign *D* scores significantly improved model fit (χ^2 = 4.19, DF = 1, p = .04), while individual differences in the Good-Bad *D* scores did not (χ^2 =

2.90, *DF* = 1, *p* = .09). The model estimate of the American-Foreign IAT (*ß* = -0.29) indicated that subjects with larger *D* scores performed poorer overall on the speech transcription task. **Figure 3** shows the relationship between overall task performance and American-Foreign *D* scores, which reflects a similar relationship as that seen in Experiment 1 (**Figure 3**).

Like Experiments 1 and 2, the main effect of counterbalance was not significant in Experiment 3 (χ^2 = 0.94, *DF* = 1, *p* = .33), but the interaction between prime and counterbalance was (χ^2 = 9.39, *DF* = 1, *p* = .002). As shown in **Figure 6B**, subjects improved from Block 1 to Block 2 in both counterbalances, but to a larger degree in Counterbalance W-EA (which was assigned to view the White face during Block 1). The interactions between the IAT measures and the effect of prime were both non-significant (*p*'s > .05).

Follow-up analyses: Block 1 data

As in Experiments 1 and 2, we also conducted analyses on data from Block 1 of Experiment 3. Our sample included 169 subjects in the East Asian prime group and 179 subjects in the White prime group. Random and fixed effects matched those in Experiments 1 and 2. Log-likelihood model comparisons are summarized in **Table 9**. Most notably, the effect of prime was not significant in the Block 1 dataset (χ^2 = 0.02, *DF* = 1, *p* = .89), indicating that the effect in the primary analysis is largely driven by Block 2 performance (see **Figure 6B**). Both the effects of the American-Foreign (χ^2 = 5.43, *DF* = 1, *p* = .02) and the Good-Bad (χ^2 = 4.83, *DF* = 1, *p* = .03) IAT *D* scores improved model fit, with trends matching the direction of those in the primary analysis of Experiment 3 (*ß*'s = -0.32 and -0.29, respectively). After accounting for outliers, the interactions between the American-Foreign and Good-Bad IAT *D* scores with the effect of prime were non-significant (both *p*'s > .05).

Table 9

Log-likelihood Model Comparisons for Block 1 of Experiment 3

Effect	$\chi^2(1)$	*p*
Prime	0.02	.89
American-Foreign *D* score	5.43	.02
Good-Bad *D* score	4.83	.03
Prime x American-Foreign *D* score	0.48	.49
Prime x Good-Bad *D* score	6.25[a]	.04[a]

Note. Values marked with superscript *a* are non-significant after removing subjects with outlying (three standard deviation from the mean) Good-Bad *D* scores.

Exploratory analyses

Bayesian analyses

As in Experiments 1 and 2, we conducted a Bayesian pair-wise t-test in JASP (JASP Team, 2023). Unlike the earlier experiments, we included the Bayesian analysis to characterize the size of the social priming effect, which was significant in the primary analyses of Experiment 3 that used GLMER. The predicted variable in this analysis was the mean proportion of keywords correctly identified by subject for each of the prime conditions (White and East Asian). Settings selected for the Student’s t-test with an alternative hypothesis of White ≠ East Asian. The default prior was used (Cauchy prior width = 0.707) and a Bayes factor robustness check was run to compare the selected prior against a wide and ultrawide prior. Results of the t-test with the default prior indicated anecdotal (weak) evidence in favor of the null hypothesis

(BF_{10} = 0.52); the wide (BF_{10} = 0.37) and ultrawide (BF_{10} = 0.27) priors from the robustness check indicated anecdotal and moderate evidence in favor of the null hypothesis, respectively.

Interim discussion

In Experiment 3, we found the first evidence within the present study that social information affects perception of speech. Specifically, transcription accuracy for the perception of Mandarin-accented English speech in noise was affected by the presentation of an East Asian versus a White prime, such that listeners performed better when presented with an East Asian prime. However, the effect of priming found in Experiment 3 was small, and would be characterized as anecdotal evidence in favor of the null hypothesis based on a Bayesian analysis (van Doorn et al., 2021). Additionally, this priming effect appeared to be driven by performance in Block 2 of the task, where a benefit was seen for subjects shown the East Asian prime as opposed to the White prime.

Of particular interest in Experiment 3 was the potential relationship between social priming effects and implicit racial associations. When examining individual differences in implicit American-Foreign and Good-Bad associations, our results indicated a general relationship between performance during the speech perception task and IAT *D* scores, but no interaction with the priming manipulation. This negative relationship showed that subjects with more negative biases (stronger associations between White faces and the construct American or Good, and between East Asian faces and the construct Foreign or Bad) also performed worse on the speech transcription task.

General Discussion

In the present study, we sought to investigate the effects of perceived race and ethnicity on perception of L1- and L2-accented English speech and explore individual listener differences.

Our primary hypothesis was that social primes (i.e., presentation of a White, East Asian, or Latina face) would affect the perception of speech presented in noise. For L1 (American) accent, we predicted a negative social priming effect, such that subjects would have poorer transcription accuracy when presented with an East Asian or Latina prime as compared to a White prime. Conversely, for L2 (Mandarin Chinese) accent, we predicted a facilitative social priming effect, such that subjects would have better transcription accuracy when presented with an East Asian prime as compared to a White prime. Based on an exemplar model of speech perception, we also predicted that listeners with larger social priming effects would have stronger implicit associations between the relevant races/ethnicities and the constructs American vs. Foreign (but not the constructs Good vs. Bad). Here, we report on the overarching trends across our three experiments.

A Lack of Social Priming for L1 Accent

Our mixed-effects analysis of social priming for perception of L1 American-accented English speech did not reveal an effect on transcription accuracy. Further, exploratory analyses using Bayesian statistics indicated strong evidence in favor of the null hypothesis. This result conflicts with the findings of Babel and Russell (2015), in which presentation of Chinese-Canadian speakers' faces resulted in poorer transcription accuracy than control (fixation cross) trials for L1-accented speech (although, see also McLaughlin & Van Engen, 2023a, in which null outcomes for L1 accent perception were also found). Several differences between the design of our study and that of Babel and Russell (2015) may have contributed to our different results: 1) Using a visual matched-guise paradigm in place of pictures of the actual speakers (which may be more believable to the participant); 2) Blocking conditions in our speech transcription task as opposed to randomly presenting both conditions intermixed (which may lessen engagement with the priming images; cf, McLaughlin & Van Engen, 2023a); and, 3) Sampling a population of White American listeners from across the United States (as opposed to focusing on a population of listeners who specifically have greater experience with the racial/ethnic group of interest).

The visual matched-guise paradigm used in the present study was adapted from the design used by McGowan (2015). This paradigm has two primary benefits: The combination of images and audio can be completely counterbalanced, and images can be matched 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 visual matched-guise paradigm in the present study 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 and Russell (2015) used stimuli from 12 L1 Canadian-accented English talkers (6 Chinese-Canadian, 6 White-Canadian), the primes were pictures of the actual speakers, and the stimuli and priming conditions were presented in a randomized order. It is possible that these design factors promote greater engagement with priming materials.

Additionally, although the present study aimed to examine social priming effects on the level of the larger population of White listeners in the United States, it is possible that these social priming effects are more likely to be observed in smaller communities with shared sociolinguistic experiences. Within our sample, some listeners may have come from communities with larger L2 speaker populations, and others smaller, resulting in different socio-phonetic mappings in the L1 listeners. Indeed, Babel and Russell (2015) discuss that in the Metro Vancouver area (from which they sampled their listeners) stereotypes of Asian Canadians as L2 speakers of English are prominent. By informing subjects in the experiment that the speakers were from Richmond (i.e., a suburb known for its Cantonese-speaking community), the authors may have strengthened their social priming manipulation. Further investigation within this community of listeners using audiovisual stimuli has also demonstrated that adaptation to L2 accent is impeded when the speaker is White (Babel, 2022), indicating that listeners anticipated the speaker to have L1 accent. In contrast, listeners were not impeded when listening to L1 accent produced by an Asian speaker. Babel interpreted this outcome as indicative of the multi-ethnic landscape of the local community: Listeners were prepared for an Asian speaker with either L1 or L2 accent, but unaccustomed and less prepared for a White speaker with L2 instead of L1 accent.

Social Priming for L2 Accent

Complementing the design of our initial experiments with L1 accents, our third experiment examined social priming for the perception of L2- (Mandarin Chinese-) accented English speech. Here, we did find evidence of social priming: Listeners had better transcription accuracy when shown an East Asian prime as compared to a White prime. While this outcome replicates prior work (McGowan, 2015; Hanulíková, 2021), it was a notably small effect; a Bayesian analysis of the data indicated that the effect should be characterized as anecdotal evidence in favor of the null hypothesis (van Doorn et al., 2021). As discussed above, it is possible that one or more of our design choices (e.g., using a visual matched-guise paradigm, blocking our task, and/or collecting data online with White subjects from across the United States) may have hindered our ability to detect a social priming effect more robustly.

However, the presence of a social priming effect for L2 accent in Experiment 3 indicates that the null results for L1 accent in Experiments 1 and 2 are unlikely attributable to an issue of methodology alone. Indeed, the designs of Experiments 1 and 3 were identical with the exception of the auditory stimuli (i.e., presentation of an L1 versus an L2 accent), and yet the East Asian and White primes only elicited differences in transcription accuracy for the experiment containing L2 accent.

Implicit Associations

In the present study, subjects in all experiments completed two types of Implicit Association Tests (IATs): One that measured associations between the races/ethnicities of interest and the constructs American vs. Foreign (“American-Foreign IAT”), and one that measured associations between the races/ethnicities of interest and the constructs Good vs. Bad (“Good-Bad IAT”). Across all experiments, the group-wide trends of these IATs were consistent: The White subjects in the present study more strongly associated White faces with the constructs American and Good, and East Asian and Latinx faces with the constructs Foreign and Bad. These results indicate an implicit group bias favoring one’s own race (in this case, White).

By including these measures of implicit associations, our primary goal was to examine how individual differences in implicit associations may predict susceptibility to social priming effects during speech perception. Further, we predicted that American-Foreign IAT *D* scores (i.e., summary statistics) would interact with social priming effects, while Good-Bad IAT *D* scores would not. More specifically, we predicted that the American-Foreign IAT may capture a unique source of variance in the social priming data because of a relationship between the construct of foreignness and expectations about L2 accent. If this were the case, the Good vs. Bad IAT would serve as discriminant validity. As we suspected, results of the models across all three experiments indicated that the two IATs did in fact capture unique sources of variance.

However, in none of the experiments did we find a relationship between either set of IAT *D* scores and social priming.

What did emerge across the three experiments was an overall relationship between IAT *D* scores and performance on the speech transcription task. Specifically, subjects with stronger American-Foreign implicit associations in the predicted direction (i.e., the construct American was more easily associated with White faces than East Asian and Latinx faces, and visa versa for the construct Foreign) had poorer overall transcription accuracy. For the Good-Bad IAT, there was a significant relationship with overall transcription accuracy in Experiment 2 (L1 accent, Latina and White primes) and for Block 1 of Experiment 3 (L2 accent, East Asian and White primes). We suggest two explanations for these trends. First, it is possible that subjects with stronger implicit associations in the predicted direction may be less engaged during the speech perception task overall due to poorer mood and/or lower motivation (i.e., in response to a speaker's perceived race or ethnicity). Note that this possibility differs from what we predicted, which was that the IAT measures could be used to examine the strength of implicit associations and thus predict the size of social priming effects. Additionally, this explanation fails to account for why IAT scores did not interact with the priming manipulation.

A second possibility is that performance on the IATs and speech transcription task may be related (at least partially) 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 (less biased) 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 (at least partially) 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 a negative relationship (as was found in all cases). While this explanation can nicely account for the American-Foreign IAT *D* scores predicting overall performance in all experiments, it does not explain why the Good-Bad IAT *D* scores only predicted overall performance in Experiment 2 and Block 1 of Experiment 3. Indeed, this divergence in outcomes for the two IATs, as well as the low variance inflation factors in our mixed-effects models, suggest that individual differences in American-Foreign and Good-Bad IAT *D* scores captured unique sources of variance.

Given the correlational nature of the IAT *D* scores with the speech transcription task, we also note another, equally plausible, interpretation of the results: Namely, it could be the case that subjects who did poorer in the transcription task (which occurred first) then had a "spill-over" effect for the IAT. This interpretation of the data is not mutually exclusive from the executive cognitive control account; indeed, participants who were more mentally depleted after the transcription task would be expected to be poorer at controlling their responses in the IAT (i.e., to avoid appearing biased). In future work, including additional measures of individual differences in executive function could help to determine whether a third variable is at play or not.

Evaluating an Exemplar Framework for Social Priming

Many authors have posited that social priming effects serve as favorable evidence for an exemplar model of speech perception (McGowan, 2015; Hanulíková, 2021; Melguy & Johnson, 2021). On this view, listeners create abstracted categories over time, effectively linking social groupings and phonetic patterns based on their unique experiences. Thus, we predicted that listeners with stronger associations between given races and ethnicities and constructs related to L2 accent (i.e., the construct “foreign”) would have larger social priming effects. Across three experiments we failed to find evidence to support this hypothesis.

Additionally, we found no evidence indicating a social priming effect occurred for L1 accent perception. Rather, the data serve as strong evidence in favor of the null hypothesis (i.e., perception of L1 accent is unaffected by the speaker’s perceived race). Based on prior key studies (e.g., McGowan, 2015; Babel & Russell, 2015), our hypothesis was that a speaker’s perceived race/ethnicity would affect the perception of both L1- and L2-accented speech, either facilitating it (i.e., an East Asian prime for Mandarin accent) or hindering it (i.e., an East Asian prime for American accent). This hypothesis was predicated on an assumption that social priming directly affects speech perception equally across listening contexts. However, the present study’s results demonstrate that the same social primes may only elicit differences in performance in the context of L2 accent. Indeed, similar results have been found for White L1 German listeners: Hanulíková (2021) found that, for teenage and older adult German listeners, priming effects for an East Asian vs. a White prime emerged for L2- (Korean-) accented German but not L1-accented German.

Our results raise the following question: Why might social priming occur in the context of L2 accent but not L1 accent, and how can an exemplar framework account for these results? One possibility is that conflicting socio-indexical and acoustic information is more easily resolved by listeners than we hypothesized. On this view, when a listener encounters what they recognize to be an L1 accent (based on acoustic input), they may inhibit socio-phonetic activation from other less-reliable sources of information. In other words, cues toward accentedness from a speaker's perceived race and ethnicity may be given less weight than acoustic cues. In contrast, when a listener encounters what they recognize to be an L2 accent, they may leverage the socio-indexical information available to them. This strategy would allow the listener to accommodate a specific type of L2 accent more effectively and efficiently. For example, in the present study, seeing an East Asian face may have allowed listeners to better identify and tune to the Mandarin Chinese L2 accent.

We note that although "negative" social priming effects were not observed in the present study, many racial and ethnic minority members experience linguistic discrimination when using their L1 in the real world. These instances of linguistic discrimination are highly consequential, affecting evaluations (e.g., of racial/ethnic minority teachers) and career opportunities. Social stigmatization is multi-faceted, and although we did not observe an effect of race/ethnicity on *perceptual accuracy* for L1 accent perception, there are other avenues through which social information may impact real-world communication (e.g., a listener's willingness to engage with a minority speaker and dedicate cognitive resources toward processing). These other effects merit further investigation in linguistic research.

Limitations and Future Directions

The most notable difference between the current study and prior work that has examined social priming effects was the method of subject recruitment and participation. Due to the COVID-19 pandemic, all experiments were designed for online, remote participation. A common concern with online data collection is that subjects may not devote their full attention to the task, as they would for an in-person study. Thus, to increase the quality of our data, we took the following precautionary steps 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 had to complete a headphone screening, and were given an additional non-penalized opportunity to report if they failed to use headphones. Subjects who demonstrated or self-reported lack of attention to the task, and/or failed to comply with headphone requirements, were excluded from analyses. Thus, we do not have reason to suspect that the quality of our data is poorer than in-lab collected data; however, it is a limitation of the current work that our methodology is less directly comparable to other studies, and particularly other studies which have garnered significant outcomes where we found null outcomes (e.g., Babel & Russell, 2015).

An additional limitation of the present study was the sample of subjects. First, 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 L1 listeners. In future work, including measures of listeners' exposure to relevant L2 accents and of the diversity of listeners' social networks may help to clarify the role of listener experience in social priming effects. Additionally, in the present study, we decided to focus on social priming effects for White listeners, based on prior evidence from our lab that indicated that social priming effects may only present in White listeners, and not non-White listeners (McLaughlin & Van Engen, 2023b). However, it will be crucial in future work to directly examine differences among listeners of different races and ethnicities. This line of inquiry can help to determine whether in-group (i.e., same race/ethnicity) preferences (see Brewer, 2007) may play a role in effects of race and ethnicity information on speech perception.

Another possible limitation of the current work is the use of transcription accuracy as the dependent measure in the speech perception tasks. To avoid ceiling effects, all the auditory stimuli in the present study were mixed with background noise, which may have reduced listeners’ abilities to identify and/or adapt to the speakers and their accents. Additionally, the inclusion of background noise to avoid ceiling effects resulted in very challenging listening conditions, with participants transcribing only approximately 57% of key words correctly across all conditions. 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 null outcomes in the present study), or whether they may be context-dependent (e.g., less easily observed in speech-in-noise).

Lastly, we suggest that future work examining individual differences in the context of social priming effects may benefit from alternatives to the IAT. A large factor in determining power in individual differences research is the precision of a given individual difference measure. Although they are well known and widely used, IATs tend to have relatively low internal reliability (e.g., $\alpha = .68$ on average in the present study) than what a rule-of-thumb would consider good ($\alpha > .70$) or ideal ($\alpha > .90$). This means that our estimate of each individual listener’s implicit associations is only moderately precise, and we therefore needed a very large sample of subjects to detect a relationship between this measure and listeners' ability to understand speech. Thus, it is possible that even with the significantly larger sample sizes present in the current study (as compared to prior work), we nonetheless had insufficient power to detect the relationship of interest. In future research, improving the reliability of IATs (or developing another measure of implicit associations with higher reliability) would provide great advantages when investigating individual differences in social priming effects.

Conclusion

A growing body of evidence indicates that listeners leverage socio-indexical information, such as a speaker's race, to process speech. In the present study, we aimed to build on this literature by not only examining effects of perceived speaker race and ethnicity on speech perception, but by also examining how implicit racial/ethnic associations may explain individual listener differences. Contrary to our predictions, evidence of social priming affecting the perception of speech was only found in the context of L2 accent, and not L1 accent. Specifically, for Mandarin Chinese- (L2-) accented English speech presented in noise, listeners had slightly better sentence transcription accuracy when presented an East Asian face than when presented a White face. These results align with an exemplar model account, but also suggest that, in the context of L1 accent perception, conflicting socio-indexical and acoustic information may be more easily resolved by listeners than we hypothesized. Ultimately, by exploring social priming effects with additional methodological approaches, and in different populations of listeners, we will be able to determine whether these effects operate differently in the context of L1- and L2-accented speech.

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.

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., 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.

Brewer, M. B. (2007). The social psychology of intergroup relations: Social categorization, ingroup bias, and outgroup prejudice. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 695–715). The Guilford Press.

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.

Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality engineering*, *14*(3), 391-403.

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.

DeJesus, J. M., Hwang, H. G., Dautel, J. B., & Kinzler, K. D. (2018). “American= English Speaker” Before “American= White”: The Development of Children's Reasoning About Nationality. *Child Development*, *89*(5), 1752-1767.

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.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt?. *Journal of Psycholinguistic Research, 38*(4), 379-412.

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.

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

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.

Hanulíková, A. (2021). Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups. *PloS one, 16*(10), e0259230.

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.

JASP Team (2023). JASP (Version 0.17.1) [Computer software].

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. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

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.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577-12580.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social cognition, 27*(4), 623-634.

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race?. *Cognition*, *119*(1), 1-9.

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.

Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, *7*(1), 1-16.

Labov, W. (1986). The social stratification of (r) in New York City department stores. In *Dialect and language variation* (pp. 304-329). Academic Press.

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

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

McLaughlin, D. J., & Van Engen, K. J. (2023a). American listeners' recognition of sentences unaffected by racial and ethnic primes. *Proceedings of the International Conference of Phonetic Sciences.*

McLaughlin, D. J., & Van Engen, K. J. (2023b). Social priming: Exploring the effects of speaker race and ethnicity on perception of second language accents. *Language and Speech.* Pre-print: https://psyarxiv.com/df2rw/

Melguy, Y. V., & Johnson, K. (2021). General adaptation to accented English: Speech intelligibility unaffected by perceived source of non-native accent. *The Journal of the Acoustical Society of America*, *149*(4), 2602-2614.

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.

Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistics Compass, 1*(5), 416-449.

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.

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.

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.

Porretta, V., & Tucker, B. V. (2019). Eyes wide open: pupillary response to a foreign accent varying in intelligibility. *Frontiers in Communication, 4,* 8.

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 Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., ... & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*, 813-826.

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.

Walker, M., Szakay, A., & Cox, F. (2019). Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?. *Laboratory Phonology, 10*(1): 7. doi: https://doi.org/10.5334/labphon.90

Wang, H., & Van Heuven, V. J. (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. *Linguistics in the Netherlands 2006*, 237-248.

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

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.

Appendix A

Experiment 1 Model Summaries

Table A1

Primary Analysis: Model with Lower Order Fixed Effects

Fixed Effects

Predictor	**Coefficient (β)**	**S.E.**	***Z***	***p***
Intercept	0.46	0.26	1.78	.07
Prime (White)	-0.02	0.36	-0.05	.96
American-Foreign *D* score	-0.24	0.12	-2.06	.04
Good-Bad *D* score	-0.06	0.10	-0.56	.57
Counterbalance (W-EA)	0.11	0.08	1.41	0.16

Random Effects

Group	**Effect**	**Variance**	**S.D.**	**Correlation**	
Subject	Intercept	0.43	0.65		
	Prime (White)	0.31	0.56	-0.30	
	Speaker (Native 2)	0.14	0.38	0.11	0.08
Item	Intercept	1.89	1.37		
	Prime (White)	0.30	0.55	0.03	
	Speaker (Native 2)	3.38	1.84	-0.60	-0.28

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score + CB +
        (1 + Prime + Speaker | Subject) + (1 + Prime + Speaker | Item),
        data = E1_alldata, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summary of the generalized linear mixed-effects regression models for the speech transcription task in Experiment 1 before adding interactions. Reference level in dummy-coding for Prime is East Asian. Reference level in dummy-coding for Counterbalance is EA-W (Counterbalance EA-W). S.E.: standard error; S.D.: standard deviation; CB2: Counterbalance 2.

Table A2

Primary Analysis: Model with Lower Order Fixed Effects and Interactions

Fixed Effects

Predictor	**Coefficient (β)**	**S.E.**	***Z***	***p***
Intercept	0.31	0.26	1.21	0.23
Prime (White)	0.32	0.36	0.87	0.38
American-Foreign *D* score	-0.19	0.13	-1.51	0.13
Good-Bad *D* score	-0.03	0.11	-0.24	0.81
Counterbalance (W-EA)	0.36	0.08	4.43	< .001
Prime (White) : American-Foreign *D* score	-0.10	0.12	-0.80	0.43
Prime (White) : Good-Bad *D* score	-0.07	0.11	-0.60	0.55
Prime (White) : Counterbalance (W-EA)	-0.58	0.08	-7.62	< .001

Random Effects

Group	**Effect**	**Variance**	**S.D.**	**Correlation**	
Subject	Intercept	0.42	0.65		
	Prime (White)	0.02	0.13	0.39	
	Speaker (Native 2)	0.34	0.58	-0.22	0.10
Item	Intercept	1.89	1.37		
	Prime (White)	0.28	0.52	0.05	
	Speaker (Native 2)	3.38	1.84	-0.60	-0.30

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score + CB +
        Prime:AF_d_score + Prime:GB_d_score + Prime:CB +
        (1 + Prime + Speaker | Subject) + (1 + Prime + Speaker | Item),
        data = E1_alldata, family = "binomial",
```

control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))

Note. Summaries of the generalized linear mixed-effects regression models for the speech transcription task in Experiment 1 after adding interactions. Rows highlighted in gray contain the lower order effects of the model, which represent partialled values and were used for interpreting the higher order effects. Reference level in dummy-coding for Prime is East Asian. Reference level in dummy-coding for Counterbalance is EA-W (Counterbalance EA-W). S.E.: standard error; S.D.: standard deviation; CB2: Counterbalance 2.

Table A3

Analysis of Block 1: Model with Lower Order Fixed Effects

Fixed Effects

Predictor	Coefficient (β)	S.E.	Z	p
Intercept	0.32	0.22	1.42	0.16
Prime (White)	0.03	0.31	0.11	0.92
American-Foreign *D* score	-0.26	0.12	-2.22	0.03
Good-Bad *D* score	-0.09	0.10	-0.83	0.41

Random Effects

Group	Effect	Variance	S.D.
Subject	Intercept	0.41	0.64
Item	Intercept	0.91	0.96

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score +
        (1 | Subject) + (1 | Item),
        data = E1_block1data, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summary of the generalized linear mixed-effects regression models for Block 1 of the speech transcription task in Experiment 1 before adding interactions. Reference level in dummy-coding for Prime is East Asian. S.E.: standard error; S.D.: standard deviation.

Table A4

Analysis of Block 1: Model with Lower Order Fixed Effects and Interactions

Fixed Effects

Predictor	**Coefficient (β)**	**S.E.**	***Z***	***p***
Intercept	0.28	0.23	1.25	0.21
Prime (White)	0.1	0.32	0.33	0.74
American-Foreign *D* score	-0.28	0.15	-1.85	0.06
Good-Bad *D* score	0.11	0.13	0.86	0.39
Prime (White) : American-Foreign *D* score	0.11	0.23	0.47	0.64
Prime (White) : Good-Bad *D* score	-0.55	0.22	-2.54	0.01

Random Effects

Group	**Effect**	**Variance**	**S.D.**
Subject	Intercept	0.40	0.63
Item	Intercept	0.91	0.96

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score +
          Prime:AF_d_score + Prime:GB_d_score +
          (1 | Subject) + (1 | Item),
          data = E1_block1data, family = "binomial",
          control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summaries of the generalized linear mixed-effects regression models for Block 1 of the speech transcription task in Experiment 1 after adding interactions. Rows highlighted in gray contain the lower order effects of the model, which represent partialled values and were used for interpreting the higher order effects. Reference level in dummy-coding for Prime is East Asian. S.E.: standard error; S.D.: standard deviation.

Appendix B

Experiment 2 Model Summaries

Table B1

Primary Analysis: Model with Lower Order Fixed Effects

Fixed Effects

Predictor	**Coefficient (β)**	**S.E.**	***Z***	***p***
Intercept	0.45	0.22	2.07	.04
Prime (White)	-0.01	0.31	-0.05	.96
American-Foreign *D* score	-0.20	0.10	-2.01	.04
Good-Bad *D* score	-0.22	0.08	-2.78	.01
Counterbalance (W-L)	0.10	0.07	1.54	0.12

Random Effects

Group	**Effect**	**Variance**	**S.D.**	**Correlation**
Subject	Intercept	0.32	0.57	
	Prime (White)	0.30	0.55	-0.39
Item	Intercept	0.91	0.95	
	Prime (White)	0.35	0.59	-0.24

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score + CB +
        (1 + Prime | Subject) + (1 + Prime | Item),
        data = E2_alldata, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summary of the generalized linear mixed-effects regression models for the speech transcription task in Experiment 2 before adding interactions. Reference level in dummy-coding for Prime is Latinx. Reference level in dummy-coding for Counterbalance is L-W (Counterbalance L-W). S.E.: standard error; S.D.: standard deviation.

Table B2

Primary Analysis: Model with Lower Order Fixed Effects and Interactions

Fixed Effects

Predictor	Coefficient (β)	S.E.	*Z*	*p*
Intercept	0.39	0.22	1.76	.08
Prime (White)	0.14	0.31	0.44	.66
American-Foreign *D* score	-0.26	0.11	-2.38	.02
Good-Bad *D* score	-0.30	0.09	-3.39	< .001
Counterbalance (W-L)	0.31	0.07	4.65	< .001
Prime (White) : American-Foreign *D* score	0.14	0.11	1.3	0.19
Prime (White) : Good-Bad *D* score	0.18	0.09	2.01	0.04
Prime (White) : Counterbalance (W-L)	-0.49	0.07	-7.34	< .001

Random Effects

Group	Effect	Variance	S.D.	Correlation
Subject	Intercept	0.31	0.56	
	Prime (White)	0.23	0.48	-0.34
Item	Intercept	0.91	0.95	
	Prime (White)	0.79	0.89	-0.42

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score + CB +
        Prime:AF_d_score + Prime:GB_d_score + Prime:CB +
        (1 + Prime | Subject) + (1 + Prime | Item),
        data = E2_alldata, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summaries of the generalized linear mixed-effects regression models for the speech transcription task in Experiment 2 after adding interactions. Rows highlighted in gray contain the lower order effects of the model, which represent partialled values and were used for interpreting the higher order effects. Reference level in dummy-coding for Prime is Latinx. Reference level in dummy-coding for Counterbalance is L-W (Counterbalance L-W). S.E.: standard error; S.D.: standard deviation.

Table B3

Analysis of Block 1: Model with Lower Order Fixed Effects

Fixed Effects

Predictor	**Coefficient (*β*)**	**S.E.**	***Z***	***p***
Intercept	0.37	0.23	1.61	.11
Prime (White)	0.04	0.32	0.11	.91
American-Foreign *D* score	-0.20	0.11	-1.89	.06
Good-Bad *D* score	-0.26	0.09	-2.97	.003

Random Effects

Group	**Effect**	**Variance**	**S.D.**
Subject	Intercept	0.30	0.54
Item	Intercept	0.97	0.99

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score +
        (1 | Subject) + (1 | Item),
        data = E2_block1data, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summary of the generalized linear mixed-effects regression models for Block 1 of the speech transcription task in Experiment 2 before adding interactions. Reference level in dummy-coding for Prime is Latinx. S.E.: standard error; S.D.: standard deviation.

Table B4

Analysis of Block 1: Model with Lower Order Fixed Effects and Interactions

Fixed Effects

Predictor	**Coefficient (*β*)**	**S.E.**	***Z***	***p***
Intercept	0.42	0.23	1.83	.07
Prime (White)	-0.07	0.33	-0.21	.83
American-Foreign *D* score	-0.35	0.15	-2.32	.02
Good-Bad *D* score	-0.28	0.12	-2.39	.02
Prime (White) : American-Foreign *D* score	0.29	0.21	1.37	.17
Prime (White) : Good-Bad *D* score	0.04	0.17	0.26	.80

Random Effects

Group	**Effect**	**Variance**	**S.D.**
Subject	Intercept	0.29	0.54
Item	Intercept	0.97	0.99

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score +
        Prime:AF_d_score + Prime:GB_d_score +
        (1 | Subject) + (1 | Item),
        data = E2_block1data, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summaries of the generalized linear mixed-effects regression models for Block 1 of the speech transcription task in Experiment 2 after adding interactions. Rows highlighted in gray contain the lower order effects of the model, which represent partialled values and were used for interpreting the higher order effects. Reference level in dummy-coding for Prime is Latinx. S.E.: standard error; S.D.: standard deviation.

Appendix C

Experiment 3 Model Summaries

Table C1

Primary Analysis: Model with Lower Order Fixed Effects

Fixed Effects

Predictor	**Coefficient (β)**	**S.E.**	***Z***	***p***
Intercept	0.61	0.17	3.61	< .001
Prime (White)	-0.09	0.04	-2.06	.04
American-Foreign *D* score	-0.29	0.14	-2.05	.04
Good-Bad *D* score	-0.23	0.13	-1.71	.09
Counterbalance (W-EA)	0.09	0.09	0.97	.33

Random Effects

Group	**Effect**	**Variance**	**S.D.**	**Correlation**	
Subject	Intercept	0.52	0.72		
	Prime (White)	0.17	0.42	0.16	
	Speaker (Native 2)	0.19	0.43	-0.27	0.15
Item	Intercept	1.34	1.16		
	Prime (White)	0.02	0.12	0.17	
	Speaker (Native 2)	2.59	1.61	-0.55	-0.17

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score + CB +
        (1 + Prime + Speaker | Subject) + (1 + Prime + Speaker | Item),
        data = E3_alldata, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summary of the generalized linear mixed-effects regression models for the speech transcription task in Experiment 3 before adding interactions. Reference level in dummy-coding for Prime is East Asian. Reference level in dummy-coding for Counterbalance is EA-W (Counterbalance EA-W). S.E.: standard error; S.D.: standard deviation.

Table C2

Primary Analysis: Model with Lower Order Fixed Effects and Interactions

Fixed Effects

Predictor	**Coefficient (β)**	**S.E.**	***Z***	***p***
Intercept	0.16	0.21	0.77	.44
Prime (White)	0.88	0.28	3.10	.002
American-Foreign *D* score	-0.24	0.14	-1.69	.09
Good-Bad *D* score	-0.23	0.14	-1.74	.08
Counterbalance (W-EA)	0.98	0.29	3.43	< .001
Prime (White) : American-Foreign *D* score	-0.22	0.13	-1.69	.09
Prime (White) : Good-Bad *D* score	0.04	0.12	0.37	.71
Prime (White) : Counterbalance (W-EA)	-1.82	0.55	-3.27	.001

Random Effects

Group	**Effect**	**Variance**	**S.D.**	**Correlation**	
Subject	Intercept	0.52	0.72		
	Prime (White)	0.15	0.38	0.21	
	Speaker (Native 2)	0.21	0.46	-0.28	0.16
Item	Intercept	1.21	1.10		
	Prime (White)	0.02	0.12	0.17	
	Speaker (Native 2)	2.59	1.61	-0.62	-0.18

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score + CB +
        Prime:AF_d_score + Prime:GB_d_score + Prime:CB +
        (1 + Prime + Speaker | Subject) + (1 + Prime + Speaker | Item),
        data = E3_alldata, family = "binomial",
```

control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))

Note. Summaries of the generalized linear mixed-effects regression models for the speech transcription task in Experiment 3 after adding interactions. Rows highlighted in gray contain the lower order effects of the model, which represent partialled values and were used for interpreting the higher order effects. Reference level in dummy-coding for Prime is East Asian. Reference level in dummy-coding for Counterbalance is EA-W (Counterbalance EA-W). S.E.: standard error; S.D.: standard deviation.

Table C3

Analysis of Block 1: Model with Lower Order Fixed Effects

Fixed Effects

Predictor	**Coefficient (*β*)**	**S.E.**	***Z***	***p***
Intercept	0.22	0.18	1.22	.22
Prime (White)	-0.01	0.08	-0.14	.89
American-Foreign *D* score	-0.32	0.14	-2.34	.02
Good-Bad *D* score	-0.29	0.13	-2.21	.03

Random Effects

Group	**Effect**	**Variance**	**S.D.**
Subject	Intercept	0.49	0.70
Item	Intercept	0.58	0.76

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score +
        (1 | Subject) + (1 | Item),
        data = E3_block1data, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summary of the generalized linear mixed-effects regression models for Block 1 of the speech transcription task in Experiment 3 before adding interactions. Reference level in dummy-coding for Prime is East Asian. S.E.: standard error; S.D.: standard deviation.

Table C4

Analysis of Block 1: Model with Lower Order Fixed Effects and Interactions

Fixed Effects

Predictor	**Coefficient (β)**	**S.E.**	***Z***	***p***
Intercept	0.23	0.19	1.21	.23
Prime (White)	-0.01	0.11	-0.07	.95
American-Foreign *D* score	-0.43	0.21	-2.07	.04
Good-Bad *D* score	-0.11	0.19	-0.60	.55
Prime (White) : American-Foreign *D* score	0.19	0.28	0.69	.49
Prime (White) : Good-Bad *D* score	-0.31	0.26	-1.21	.23

Random Effects

Group	**Effect**	**Variance**	**S.D.**
Subject	Intercept	0.49	0.70
Item	Intercept	0.58	0.76

R Model Syntax

```
glmer(cbind(Correct, Incorrect) ~ 1 + Prime + AF_d_score + GB_d_score +
        Prime:AF_d_score + Prime:GB_d_score +
        (1 | Subject) + (1 | Item),
        data = E3_block1data, family = "binomial",
        control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

Note. Summaries of the generalized linear mixed-effects regression models for Block 1 of the speech transcription task in Experiment 3 after adding interactions. Rows highlighted in gray contain the lower order effects of the model, which represent partialled values and were used for interpreting the higher order effects. Reference level in dummy-coding for Prime is East Asian. S.E.: standard error; S.D.: standard deviation.