ACCENT BIAS AND RACIAL ACCENTS IN ACADEMIC COMPREHENSION

BY

VICTORIA SUSBERRY

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Educational Psychology
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Adviser:

Professor, Kiel Christianson

ABSTRACT

The moment a person speaks, listeners are given access to a wealth of information: class, race, gender, and other environmental and cultural details. When we listen to people speak, we are checking to see if we can understand what is being said. Accents, however, can cause someone to deem another person unintelligible not because of the content of their production, but rather because of how they said it—the phonological or acoustic properties of the production. This study seeks to determine if racial accents are perceptible in quasi-academic (academic level, but not in an academic setting) contexts, and how much visual cues to a speaker's race influence how or whether they are heard and the comprehension of the information conveyed in the various accents. To test these hypotheses, three online experiments consisting of audio recordings and photos were administered to a total of 240 participants split into three groups. In all three groups, participants were presented with photos of three women of different races (white, Latina, Black) who were depicted as having recorded the texts played auditorily that the participants read. In each experiment, two of the voices and photos were "mismatched": the speaker's voice and photo depicting the race of the speakers were swapped. In each experiment, then, one photo-voice pairing was the "matched" baseline. This design, over the course of all three experiments, attempted to dissociate audio and visual input in determining whether and how both factors might contribute to both comprehension of texts and affective responses to the speakers. The results show clear evidence of accent bias, though the unique contributions of pictures and voices remain somewhat unclear.

ACKNOWLEDGEMENTS

This thesis would be not be possible without collaboration, support, and guidance from my advisor Kiel Christianson, and my colleague, Jack Dempsey from the Department of Educational Psychology at The University of Illinois at Urbana-Champaign. I am thankful for the support of my parents, friends, and family. I would also like to thank and honor Dr. David Zola, who convinced me that I had what it takes to participate in graduate study and for believing in me when I could not do so myself.

FOREWORD

While I have yet to become an educator in the most literal sense, I have experienced the struggles and witnessed the chagrin that comes as a result of being the “other” in the American educational system. I know what it is to be discredited and devalued. I’ve gotten the back-handed compliment of “speaking well” as if I too, did not grow up in America. As if English was not also my first language. As if my linguistic capabilities surprised them, despite the circumstance of attending one of the top public universities in this country. I have never bought the lie that the way someone speaks is an indicator of their intelligence. I have heard some of the most unique colloquialisms flow from the mouths of doctors, educators, and professors, who have all done brilliant work without having to sacrifice the speech that makes them who they are. I want it to be understood that my identity, who I am as a person, is deeply connected to this research. There is no part of this work that is not intertwined with my inner self. The presented scholarship comes from a sense of righteous indignation and a personal obligation to social justice and the destruction of racial and gender inequality within education and academia specifically. By studying the racial implications of accent bias in the university setting, it is my hope to inform and support efforts to develop culturally and linguistically accurate instruments for evaluating women (and men) of color within academia—instruments that take into account biases against “nonstandard” varieties of English and people who speak them.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

Racial Linguistics

Linguistic prejudice is defined as a form of prejudice in which people hold implicit biases about others based on the way they speak (Yao, 2018). In America, there have been many instances of linguistic prejudice, such as the 2012 Trayvon Martin case. Martin was on the phone with his friend Rachel Jeantel during the time of the assault that ended in his death at the hands of George Zimmerman. Although her perspective on the incident should have made her a valuable witness on behalf of the Martin family, her speech operated against her. The jurors found her “hard to understand” and deemed her “not credible.” In the 16 hours that the jury deliberated, her testimony was not mentioned once (King & Kinzler, 2020).

Previous research done by linguist John Baugh (1999), who coined the phrase “linguistic profiling,” has demonstrated that people make different choices over the phone when they think the person they are speaking to is white versus Black. Baugh called potential landlords and changed his voice when interacting with them to see if they would allow him to view apartments. He called each landlord three separate times using either an African American Vernacular English accent, a Standard American English accent, or a Chicano English accent. He was more likely to receive a “no” if he the landlord perceived him to be Black or Chicano and much more likely to be told “yes” when they believed his voice to be attached to a white man. Similarly, economist Jeffrey Grogger (2009) found that a person’s accent or dialect is linked to specific wage inequalities. He found that among Black people, speech patterns were directly associated with the amount of school they had attended and their scores on the Armed Forces Qualification Test (AFQT). Additionally, these same speech patterns were linked to their wages. He controlled for skill level and family background, yet Black speakers who could be distinctly identified as Black were paid about 12% less than their white counterparts with comparable skills. Black speakers whose speech was not immediately associated with being Black were paid similarly to other white workers. Both of these studies highlight two very harmful instances in American society where people's racial identity negatively impacts their lives and the lives of others through their speech alone.

Accent Bias

The moment a person speaks, listeners gain access to a wealth of information: class, race, gender, and other environmental and cultural details. Listeners in any communicative context must continually monitor to ensure that language input can be understood. Accents, however, can cause a listener to deem another person unintelligible not because of what the speaker said or did not say, but rather how they said it. In a study on the effects of foreign accents on employment-related decisions, Hosoda and Stone-Romero (2010) suggest that people can still have negative judgments towards those with nonnative (differing) accents even if the person's speech is understood. In another study examining customer service calls and employee's accents (Wang et al, 2012), results show that the ability to suppress negative biases requires both a cognitive and an affective effort on the listener's part. This implies that the mood of the listener could also inhibit or allow listeners to display or hide their biases. Shah (2019) conducted a study of the psychosocial, behavioral, and personality attributes of seven common accents using voice recordings and a one through seven Likert scale. Shah found that most underlying judgements are created through media. Aside from character traits, many participants inferred occupational positions of the people they heard speak. She also found that harsh accent judgement was not reserved solely for foreign accents but regional ones as well. These results show that when people make accent judgements, biases exist not solely for foreign accents, but also for regional accents, which can be subject to both biases and stereotypes. Shah's study provides insight into understanding how others judge accents and supports the notion of stereotype influence when people hear accents that diverge from their own. In an online instructional setting, Sanchez and Khan (2016) engaged in accent research examining spoken presentations and whether or not instructor accents affected student's learning experience by listening to audio recordings of instructors teaching identical lessons and then being rated in a post test. They found that although having accented instruction does not harm or inhibit learning in any way, it causes students to view the instructor as less effective. This discovery is particularly harmful if we think of the evaluation forms at the end of the semester at universities, in which students provide feedback on the class and the instructor. If these findings can be applied generally, they suggest that the instructional efficacy of professors and instructors that ascribe to culture, environment, race, ethnicity, etc. that is not white and/or American may be evaluated unfairly by their institutions, too. Bestelmeyer, Belin, and Ladd (2015) found a heightened sensitivity to in-group speakers when they had participants listen to accented speech from three separate speakers. This result is not necessarily evidence of accent bias, but it suggests that people are more comfortable with the speech of someone who presents as similar to themselves, which could lead to a bias about other-accented speech.

Racism and Language

It is imperative that the current research be situated within two scopes. The first and most important being that (1) this research is centered around the expression of racism within contexts of interpersonal communication, and (2) it strives to understand language through the lens of the systemic racism that is alive and well in America. In August of 2020, the Conference on College Composition and Communication (CCCC, 2020) penned an open letter demanding Black linguistic justice. They asked that teachers stop requiring (Black) students to code-switch and instead teach students about the ways that white supremacy is furthered through linguistics. Code-switching in the academic setting implies that the native speech patterns are inferior or substandard to the mainstream. They asked that teachers develop and teach Black linguistic consciousness to their students to help them unlearn and unravel the ways in which racism has impacted their views on language. These demands are only two of several examples of how racism in academia, specifically racism through language expectations, is a block for non-white students. And while this piece is specific to Black students, it can be applied across cultures. According to the Merriam-Webster Dictionary (2020), standard English is "the English that with respect to spelling, grammar, pronunciation, and vocabulary is substantially uniform though not devoid of regional differences, that is well established by usage in the formal and informal speech and writing of the educated, and that is widely recognized as acceptable wherever English is spoken and understood." Looking at this definition through a critical lens, it states that Standard English is the speech and the writing of the educated, thereby implying that people who do not use this form of English are uneducated. This, however, is the not angle this research focuses on. The present research goes deeper to ask if people of color assimilate to this expectation in an effort to be seen as educated, will they still be judged based on racial associations with their accents? My main research questions are as follow. Are racial accents perceptible in quasi-academic contexts? How much do visual cues to a speaker's race influence how or whether they are heard and the comprehension of the information conveyed in their respective accents?

CHAPTER TWO: THE CURRENT STUDY

To address these questions, a paradigm was developed in which quasi-academic texts such as in Fig. 1 were selected from various sources.

> Named *Ferrodraco lentoni,* the new fossil is far from a full skeleton; it includes parts of the upper and lower jaw, five partial neck bones, sections of both wings, and many teeth. Despite pterosaur finds in Australia being exceedingly rare, these fossils are exceptionally well preserved. Described today in *Scientific Reports*, the newly found pterosaur is roughly 96 million years old, based on previously reported ages of the rock formation thought to entomb the creature's bones. Its closest relatives, pterosaurs of the group Anhangueria (an-hang-GWEHR-ee-ah), are thought to have died out by 94 million years ago. Though its precise age remains uncertain, *Ferrodraco* joins a number of exciting new Australian discoveries from the age of dinosaurs, including the most complete dinosaur fossil preserved as opal.

Figure 1. Example text used in the experiments (adapted from HISTORY.com)

The texts were then recorded by speakers of different races (see Chapter 3 for methods). Recordings were paired with stock photos of three women who matched the race of the speaker to guarantee the correct effect while maintaining the speakers' anonymity. In each of the three experiments reported below, two speakers and photos were crossed, such that in each experiment, there was one photo-voice pair that "matched," and two that "mismatched."

The working hypothesis, based on the work of Tajfel and Turner's (1979) social identity theory stages that serve to explain how people determine in-groups versus out-groups, was that participants would generally react more positively to both same-group photos and voices/accents. This hypothesis stems from stages two and three of Tajfel and Turner's model, which states that in the Social Identification stage we take the identity of the group that we have determined we belong to, meaning that our participants will have had to determine a group they belong to, which is deduced when we ask them to state their race. Then, in the Social Comparison stage once we have identified ourselves and the group that we belong to, we tend to compare the group we are in to other groups, meaning participants will be comparing the speakers to the one who most matches their self-identified race. Of course, the latter prediction is dependent on the ability of participants to discern differences between the accents of the three speakers.

An alternative hypothesis is that participants might not disfavor any particular photo or accent, but rather would react negatively to “mismatches” between photos and accents. This hypothesis is derived from previous work that reports negative sociolinguistic reactions to Black English speakers who “sound too white” or white English speakers who “sound black” (Baugh, 2018). In a study about immigrant Africans, Baugh found that non-native Black speakers of Academic English were less likely to be accepted in social settings than their white American counterparts (Smith, 2020). They also reported that the ways in which we distinguish Black speech in academia furthers feelings of inferiority and illegitimacy in Black Americans no matter how well they use English because their use of it, “perfect” or “imperfect,” is not accepted in the way that White Americans are. This is evidence that non-native and native speakers of a language who speak that language “too well” are mistrusted or not liked if they do not fit expectations associated with that speech.

CHAPTER THREE: EXPERIMENTS

Experiment 1

Building on previous research on accent bias in academia, the present research acknowledges that accent bias is real and seeks to determine if racial accents are perceptible in quasi-academic contexts, and how much visual cues to race influence how or whether they are heard and the comprehension of the information conveyed in the various accents. Specifically, I examine both the visual impact of a photograph of the "speaker" and the auditory impact of the voices of speakers of three racial accents: Black, Latina, and White. In Experiments 1, 2, and 3, the audios and pictures of three women (Black, Latina, and White) are crossed iteratively across three experiments. The goal was to determine whether accent, photo, or the matching/mismatching of photo and accent cues would influence either comprehension of the text or affective response to the speakers, or both.

Method

Participants

Eighty participants were recruited through Amazon's Mechanical Turk and were paid $7 for their participation in this study. All participants were native English speakers living in the United States with a HIT approval rating of 95% or higher. Participants were not allowed to participate in Experiments 2 or 3.

Materials

We adapted 10 excerpts of articles about varying topics in an effort to keep participants engaged while also providing them with information that could be quickly learned. (All texts are provided in full, with attributions, in the Appendix.) Each of the ten texts were reduced to 100-150 words and controlled for readability score via the Automated Readability Index (ARI, https://readabilityformulas.com/free-readability-formula-tests.php). Mean length of texts was 127 words (*SD* = 25.6), and Mean readability score (grade-level) was 15.3 (SD= 0.3). Finally, comprehension questions were generated that were meant to be simple but that required participants to listen if they were to answer correctly. Each text was followed by two questions, as seen in Fig. 2. Predictions about comprehension accuracy differed, depending on the theory one adopts. If comprehension is influenced by in-group/out-group status of speaker and hearer (Tajfel & Turner, 1979), it is expected that texts spoken by speakers sharing the same racial identity with the participant will be comprehended better. If this comprehension is due more to visual identity than accent identity, then "mismatched" photos will display similar patterns to the "matched" photos. If, however, speakers who do not sound close enough to their perceived racial accent are not preferred, a mismatch penalty should be observed, irrespective of racial identities of speakers and participants.

In his 1875 writing, *Critique of the Gotha Program*, Karl Marx summarized the communist philosophy in this way: "From each according to his ability, to each according to his needs." By contrast, socialism is based on the idea that people will be compensated based on their level of individual contribution to the economy. Unlike in communism, a socialist economic system rewards individual effort and innovation. Social democracy, the most common form of modern socialism, focuses on achieving social reforms and redistribution of wealth through democratic processes, and can co-exist alongside a free-market capitalist economy. (15.6, ARI)

1. Socialism is based on the idea that people will be compensated based on their level of individual contribution to the economy. (Y/N)

2. Social democracy cannot co-exist alongside a free-market capitalist economy. (T/F)

Figure 2: Sample Text and Questions

Three stock photos of young adult women were taken from the internet. One photo was of a Black woman, one of a Latina woman, and one of a White woman. The subject of each photo had a similar visual background, had a similar expression, and wore similar clothing. We specifically wanted the women to appear kind and inviting, thus all three photos were of smiling, casually dressed women (Fig. 3)

Figure 3. Black "speaker," Latina "speaker," White "speaker"

Once the texts had been adapted and selected, three American women from different racial backgrounds with Midwestern American English as their first language recorded the audio stimuli. Speakers were chosen according to the following rationale: if accents are distinct to race, we should be able to choose any person from any given race and be able to determine their race based on how they sound. As a result, I chose three women from my class/lab who were of the target races we wanted to measure for the experiment. The speakers were not chosen because they spoke with especially distinct or strong accents. In fact, the idea was for each speaker to sound as similar as possible to the other without losing their own speech patterns. It should also be noted that all three women are involved/pursuing careers in academia. Each woman was then recorded reading the selected texts as we measured for pronunciation, timing, volume, and enunciation. Each text recording lasted approximately 60 seconds. Recordings were not required to be flawless; minor pauses or self-corrected disfluencies were allowed, as long as they did not hinder global fluency or significantly increase the duration of the recording. Once the audios were recorded, they were then uploaded onto PRAAT (Boersma &Weenink, 2021), an audio editor program, and were stabilized and matched in pitch to ensure a similar listening experience for all three speakers.

A speaker attitude survey was created based on the "Doll Test" by Clark and Clark (1947), which was an experiment to determine racial preference in children. They asked questions such as "Which doll is prettiest?" and "Which doll is meanest?" As this instrument seemed to be profoundly efficient in determining the prejudices held by their participants, the current survey was modeled on it (Figure 4). Participants could only make one selection per question. It was predicted that participants would be more likely to select the positive options for the race nearest to them, and negative options for the race furthest from them. Specifically, we had 9 survey questions, 4 of which had a positive valence and four that had a negative valence. Question number 3 stood alone as its own measure, however it did not yield significant results and is not reported in this study.

In Experiment 1, the "matched" voice and photo set was White. The voice of the Latina speaker was paired with the Black photo, and the voice of the Black speaker was paired with the Latina photo (see Table 1).

1. Which speaker was most comprehensible?
2. Which speaker was least comprehensible?
3. Which speaker had the strongest accent?
4. Which speaker seemed more educated?
5. Which speaker seemed more trustworthy?
6. Which speaker seemed least trustworthy?
7. Which speaker seemed least educated?
8. Which speaker seemed most likeable?
9. Which speaker seemed least likeable?

Figure 4: Sample Survey Questions

Procedure

Participants were recruited through Amazon's Mechanical Turk and were able to complete the experiment through the Ibex Farm server (Drummond, 2013) With this being an online server, participants were not monitored as they completed the experiment. Each participant was required to consent electronically by typing their Amazon Mechanical Turk ID into our consent form. Once consent was given, they moved to the instructions screen, which explained that this was a comprehension experiment where they would listen to audio and answer questions after each recording. It also mentioned that there would be survey questions at the end of the experiment. Participants had to listen to the full audio before an arrow would appear that would allow them to go to the next page, which held the first question with the options of either Y/N or T/F. They would answer the question, and it would automatically move to the second question. Then the next audio would play and the process repeated itself for a total of ten audio files and twenty questions. Once this is finished participants were then asked to provide their age range, race, gender, and subject ID (In experiment 3, we also asked for their state to see if location influenced the biases but it proved non-significant is not included in this discussion of the study). Finally, participants were then taken to the ten-question survey. For each question, they saw all three photos with a selection bubble beneath and N/A if they could not make a choice. In response to each question, participants were instructed to choose one photo. There was no time limit for their responses, and response times were not recorded.

Results

Data were cleaned based on two separate measures: comprehension accuracy and survey responses. Participants with low mean accuracy scores (< 50%) on the comprehension questions were excluded, resulting in the removal of eight participants. We also removed participants who either put N/A for all their responses or selected the same answer for all nine survey questions, resulting in removal of five more participants, leaving us with a total of 68 participants in

Experiment 1. We also removed comprehension response items whose reaction times were greater than 2.5 SD away from the grand mean, resulting in the loss of one data point.

Accuracy Analysis

For the comprehension question accuracy analysis, a logistic mixed effects model was fit to the data with the lme4 package (Bates et al., 2015) in R version 1.4.1103. (R Core Team, 2020). We started by revaluing each question by the race of the woman who spoke. We did this to see if accent bias and/or mismatching auditory and visual cues induced cognitive processing load (lower accuracy results), in addition to whether participants did better or worse with the voice from their own self-identified race, despite any mismatch. Comprehension items are reported by condition in Table 1. The model included fixed effects of Participant Race and Speaker Race as well as their interaction. These fixed effects were sum coded for valence (-.5 and .5) and then treatment coded with Other being the baseline for both fixed effects as it is our control across experiments. The model is shown below.

Regression equation: Accuracy ~ Participant Race * Speaker + (1|Participant) + (1|Item)

Audio Number	Race of Woman Speaking	Picture Displayed
R1	White	White
R2	White	White
R3	White	White
R4	White	White
R5	Black	Latina
R6	Black	Latina
R7	Black	Latina
R8	Latina	Black
R9	Latina	Black
R10	Latina	Black

Table 1:. Audio/Visual pairings for Exp. 1

Results from the accuracy analysis are reported in Figure 5 and Table 2. White participants had overall higher accuracy scores than other participants (est.= 0.574, z =2.53, p = .0115). We also found that Black participants performed with higher accuracy with the Latina speaker (paired with the Black photo) (est. =0.047, z = .29, p = .772) and the Black speaker (paired with the Latina photo) (est. = 0.614, z = 3.637, p = .0002). Overall, it appears that each group had higher accuracy with the speaker of the group most nearest their own identity, supporting the social identity hypothesis (Tajfel & Turner, 1979).

	Estimate	Std. E	z value	p value	
White Participants	0.57411	0.22724	2.526	0.0115	*
Black Participants	-0.08115	0.27505	-0.295	0.767	
Latina Speaker	0.09939	0.54449	0.183	0.853	
Black Speaker	-0.13562	0.54486	-0.249	0.803	
White Participant * Latina Speaker	-0.4242	0.14073	-3.014	0.002	**
Black Participant * Latina Speaker	0.04771	0.16477	0.29	0.772	
White Participant * Black Speaker	-0.03512	0.1426	-0.246	0.805	
Black Participant * Black Speaker	0.61462	0.16899	3.637	0.0002	***

Table 2. Experiment 1 Accuracy model output

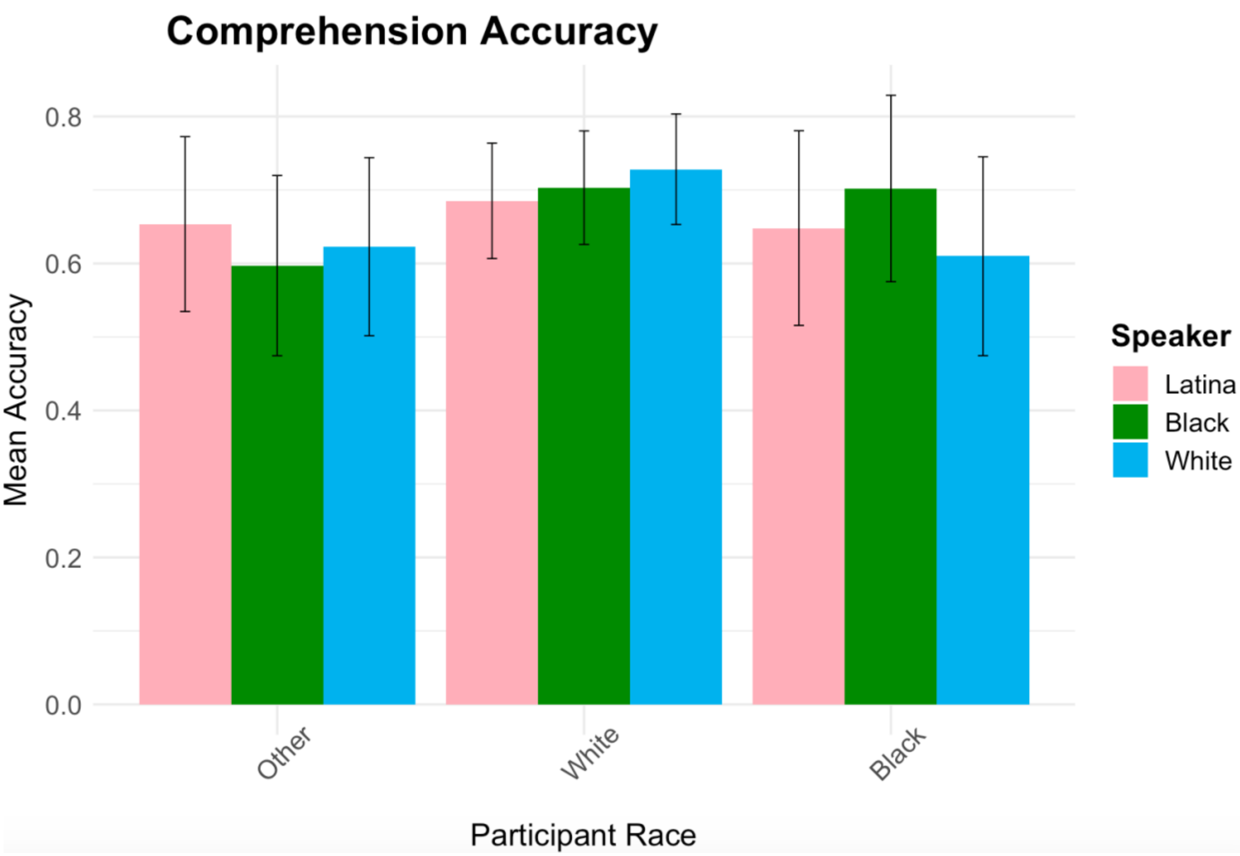

Figure 5: Exp. 1 Accuracy Means Comparing Participant Race with Speaker Race

Survey Analysis

To complete the survey analysis, a logit mixed effect model failed to converge, so logistic regressions without random slopes were fit to the data. These models contained random intercepts, but if the model failed to converge, the random intercept that was least useful was removed. If the model failed to converged with a single random intercept, we then ran a logistic regression without random effects. Survey questions were categorized according to whether or not they had a positive or negative valence (if the question asked for the “least” or the “most,” positive attribute vs. negative). Of the survey questions, approximately half were positive and half were negative. The breakdown of survey items and their valence is reported in Table 3.

Positive Valence	1. Which speaker was most comprehensible? 2. Which speaker seemed more educated? 3. Which speaker seemed more trustworthy? 4. Which speaker seemed most likeable?
Negative Valence	• Which speaker was least comprehensible? • Which speaker seemed least trustworthy? • Which speaker seemed least educated? • Which speaker seemed least likeable?

Table 3: Valence of survey questions

Three separate regression models were run for each race. There was not a large enough Latina/o representation for their own model, so the “Other” category was created here and in experiments 2 and 3. The other category included Latino/a, Asian, and Native Americans. A full accounting of the demographics of all participants in all three experiments can be found in the Appendix.

The results displayed in Table 4 provide the results of the three regression models run in this experiment on the survey data. Each panel of Table 4 displays how each of the three groups responded in the survey. Figure 6 shows participant responses to the survey questions. The first row shows how white participants responded for positive-valence questions vs. negative-valence questions. The second row shows how Black participants responded for positive-valence questions vs. negative-valence questions. The third row shows how Other races responded for positive-valence questions vs. negative-valence questions. The x-axis on each panel represents the photos the participants were viewing and choosing from when they were responding to each question. The y-axis represents the percentage of total responses given by choosing the given photo. Percentages within each panel add up to 100%.

Looking at the regression model results in Table 4 for the White responses (i.e., when participants chose the White photo for their answers), illustrated in the middle two panels labeled "Black Positive" and "Black Negative" of Figure 6, there is a simple effect of participant race such that Black participants are significantly different than the baseline group (Other) demonstrating that they were less likely to choose the White speaker for survey responses overall (est. = -.154, t = -2.435, p = .015). In the regression model for the Black responses (i.e., when the participants chose the Black photo for their answers), a significant interaction emerged between White participants and the Black photo, showing that they chose the Black photo more often when presented with negative attributes (est. = -0.29, t = -2.746, p = .006). In the regression model labeled "Other Response" (i.e., when the participants chose the Latina photo for their answers) there was a significant interaction such that the White photo (est. = .392, t = 3.691, p = .0002) and the Black photo (est. = .262, t = 2.065, p = .0395) were chosen more often for positive valence questions. Additionally, there is a simple effect of participant race such that Black participants are significantly different than the baseline in that they chose the Latina photo for more positive valence questions, which can also be seen in Figure 6 in the middle panels (est.= .205, t = .063, p = .001). These results are summarized in Table 4.

White Response						
		Estimate	Std. Error	t value	Pr (>|t|)	
	Valence	0.157	0.086	1.827	0.068	
	Participant Race White	-0.077	0.053	-1.449	0.148	
	Participant Race Black	-0.154	0.063	-2.435	0.015	*
	Valence: Participant Race White	-0.102	0.106	-0.957	0.338	
	Valence: Participant Race Black	-0.201	0.127	-1.587	0.113	
Black Response						
		Estimate	Std. Error	t value	Pr (>|t|)	
	Valence	-0.032	0.086	-0.37	0.711	
	Participant Race White	-0.027	0.053	-0.51	0.611	
	Participant Race Black	-0.05	0.063	-0.799	0.424	
	Valence: Participant Race White	-0.29	0.105	-2.746	0.006	**
	Valence: Participant Race Black	-0.061	0.126	-0.48	0.631	
Other Response						
		Estimate	Std. Error	t value	Pr (>|t|)	
	Valence	-0.125	0.086	-1.457	0.146	
	Participant Race White	0.103	0.053	1.956	0.051	
	Participant Race Black	0.205	0.063	3.23	0.001	**
	Valence: Participant Race White	0.392	0.106	3.691	0.0002	***
	Valence: Participant Race Black	0.262	0.123	2.065	0.0395	*

Table 4: Experiment 1 Survey Model

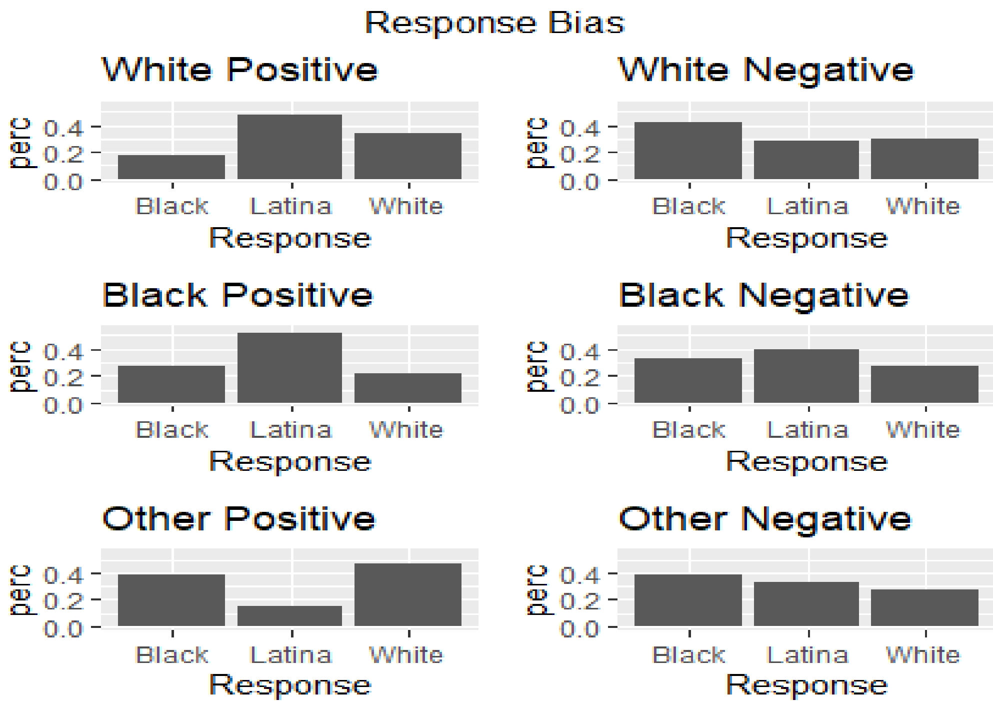

Figure 6: Survey Response Model Exp. 1: This figure shows participant responses to the survey. In the first row you see how white participants chose for positive questions v. negative questions. In the second row you see how Black participants chose for positive questions and negative questions. Finally, in the third row you see how other races chose for positive v. negative responses. It should be noted that participants were looking at the photos when they were responding.

Discussion

Based on the accuracy analysis results, it appears that participants were more likely to answer the comprehension questions correctly if they were listening to a speaker of their own race, irrespective of whether the voice “matched” or “mismatched” the photo. This result suggests that participants were sensitive to the accent of the speaker, regardless of what the photo on the screen was while they were listening.

The comprehension accuracy results contradict survey results, however. Looking at the first row in Figure 7, white participants chose the Black photo for more negative valence questions even though their accuracy results did not differ significantly between the Black and white speaker. This could be evidence of an implicit visual bias, as the survey showed only the photos and not the audio. Additionally, Black participants seemed more inclined to choose nonwhite responses no matter what was being asked, which again might imply implicit bias.

Overall, Experiment 1 yielded evidence of potential accent and visual biases. Participants exhibited better comprehension accuracy when listening to their racial in-group voice, but were more negative towards their out-group race photos in their survey responses.

Experiment 2

The methods and procedure are the same in Experiment 2 as in Experiment 1 except that in Experiment 2, the Black photo was matched with the audio of the Black speaker, whilst the White and Latina photos and audios were crossed.

Results

Data Cleaning

Data were cleaned based on two separate measures, the first being accuracy and the second being survey responses. Participants with low mean accuracy scores (< 50%; N=10) on the comprehension questions were excluded. We then further removed participants (N=2) who either put N/A for all their responses or selected the same answer for all nine survey questions. Data from the remaining 67 participants were included in data analyses.

Accuracy Analysis

The same methods used in Experiment 2 for the comprehension question accuracy analysis as in Experiment 1. The model included fixed effects of Participant Race and Speaker Race as well as their interaction. These fixed effects were sum coded for valence (-.5 and .5) and then were treatment coded with Other being the baseline for both fixed effects as it is our control across experiments. The model is shown below.

Regression Equation: Accuracy ~ Participant Race * Speaker + (1|Participant) + (1|Item)

Audio Number	**Race of Woman Speaking**	**Picture Displayed**
R1	Black	Black
R2	Black	Black
R3	Black	Black
R4	Black	Black
R5	Latina	White
R6	Latina	White
R7	Latina	White
R8	White	Latina
R9	White	Latina
R10	White	Latina

Table 5(cont): Audio/Visual pairings for Exp.2

Results from the accuracy analysis are reported in Figure 8 and Table 6. Black participants had overall lower accuracy scores than other participants (est. = -0.686, z = -2.133, p = 0.033), irrespective of speaker voice. The only other significant result found in the accuracy is that Black participants performed significantly better with the White speaker than the other two represented races (est. = 0.623, z = 3.692, p = < .0002)

	Estimate	Std. E	z value	p value	
White Participants	0.08505	0.28479	0.299	0.765221	
Black Participants	-0.68669	0.32194	-2.133	0.032927	*
White Speaker	-0.22764	0.54851	-0.415	0.678082	
Latina Speaker	-0.1832	0.55115	-0.332	0.739595	
White Participant * White Speaker	0.0859	0.15088	0.569	0.569121	
Black Participant * White Speaker	0.62301	0.16876	3.692	0.000223	***
White Participant * Latina Speaker	-0.26122	0.1538	-1.698	0.089434	
Black Participant * Latina Speaker	0.17652	0.17152	1.029	0.303393	

Table 6:Experiment 2 Accuracy Model

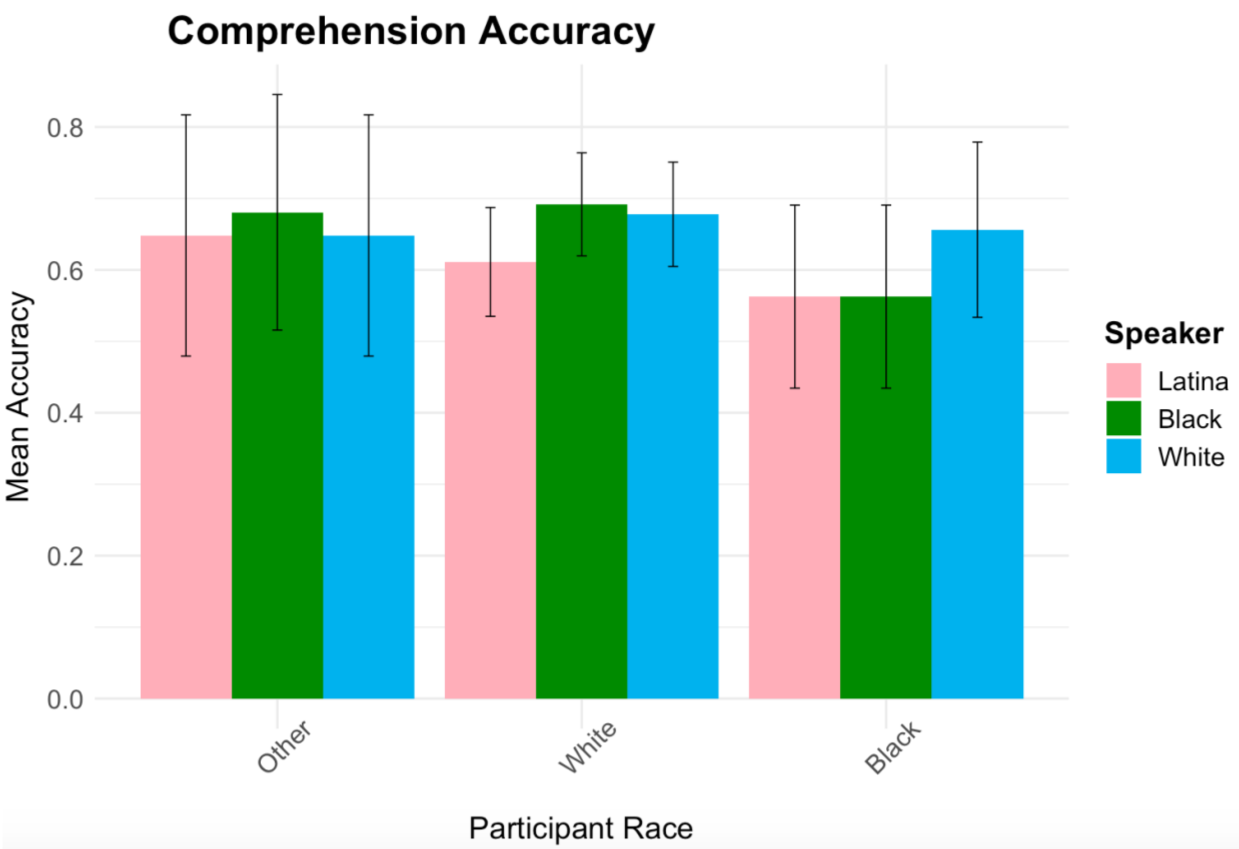

Figure 7: Accuracy Means Comparing Participant Race with Speaker Race

Survey Analysis

Similarly to experiment one, three separate regression models were run for each race. There was not a large enough Latina/o representation for their own model, so the “Other” category was created here.

The results displayed in Table 7 provide the results of the three regression models run in this experiment on the survey data. Each panel of Table 7 displays how each of the three groups responded in the survey. Figure 8 shows participant responses to the survey questions. The first row shows how white participants responded for positive-valence questions vs. negative-valence questions. The second row shows how Black participants responded for positive-valence questions vs. negative-valence questions. The third row shows how Other races responded for positive-valence questions vs. negative-valence questions. The x-axis on each panel represents the photos the participants were viewing and choosing from when they were responding to each question. The y-axis represents the percentage of total responses given by choosing the given photo. Percentages within each panel add up to 100%.

Table 7 shows in the first regression model "White Response" (i.e., when participants chose the White photo for their answers) there was a main effect of valence for the response "White" for negative survey questions overall. (est. = -2.454, z = -2.992, p = <.002), this can also be seen in the bars labeled "White" on the panels on the right side of figure 8. In the second regression "Black Response" (i.e., when the participants chose the Black photo for their answers) there was a main effect of valence for the response "Black" for positive survey questions overall (est. =0.563, t =4.886, p =1.37e-06) which can also be seen in figure 8 in all three of the panels on the left side.. Additionally, White (est. = -0.4724, t = -3.759, p = <.0001) and Black participants (est. = -0.296, t = 0.141, p = .035) had a significant interaction with the Black photo for positive attributes in comparison to the baseline group (Other). In the third regression model "Other Response" (i.e., when the participants chose the Latina photo for their answers) White participants had a significant interaction with the Latina picture for positive attributes (est. = 0.33414, t = 2.627, p = <.008). This can be seen in Table 7 and Figure 8.

White Response						
		Estimate	Std. Error	z value	Pr (>|t|)	
	Valence	-2.454	0.8202	-2.992	0.002771	**
	Participant Race White	0.7134	0.4251	1.678	0.093301	
	Participant Race Black	0.7538	0.4512	1.671	0.094793	
	Valence: Participant Race White	1.337	0.8502	1.572	0.115843	
	Valence: Participant Race Black	1.5656	0.9025	1.735	0.082806	
Black Response						
		Estimate	Std. Error	t value	Pr (>|t|)	
	Valence	0.5625	0.11513	4.886	1.37E-06	***
	Participant Race White	-0.11312	0.06283	-1.8	0.07236	
	Participant Race Black	-0.06183	0.07049	-0.877	0.380812	
	Valence: Participant Race White	-0.4724	0.12566	-3.759	0.000189	***
	Valence: Participant Race Black	-0.29688	0.14099	-2.106	0.035706	*
Other Response						
		Estimate	Std. Error	t value	Pr (>|t|)	
	Valence	-0.1875	0.11654	-1.609	0.10826	
	Participant Race White	0.03456	0.0636	0.543	0.58707	
	Participant Race Black	-0.0211	0.07136	-0.296	0.76764	
	Valence: Participant Race White	0.33414	0.1272	2.627	0.00887	**
	Valence: Participant Race Black	0.11317	0.14273	0.793	0.4282	

Table 7::Experiment 2 Survey Model

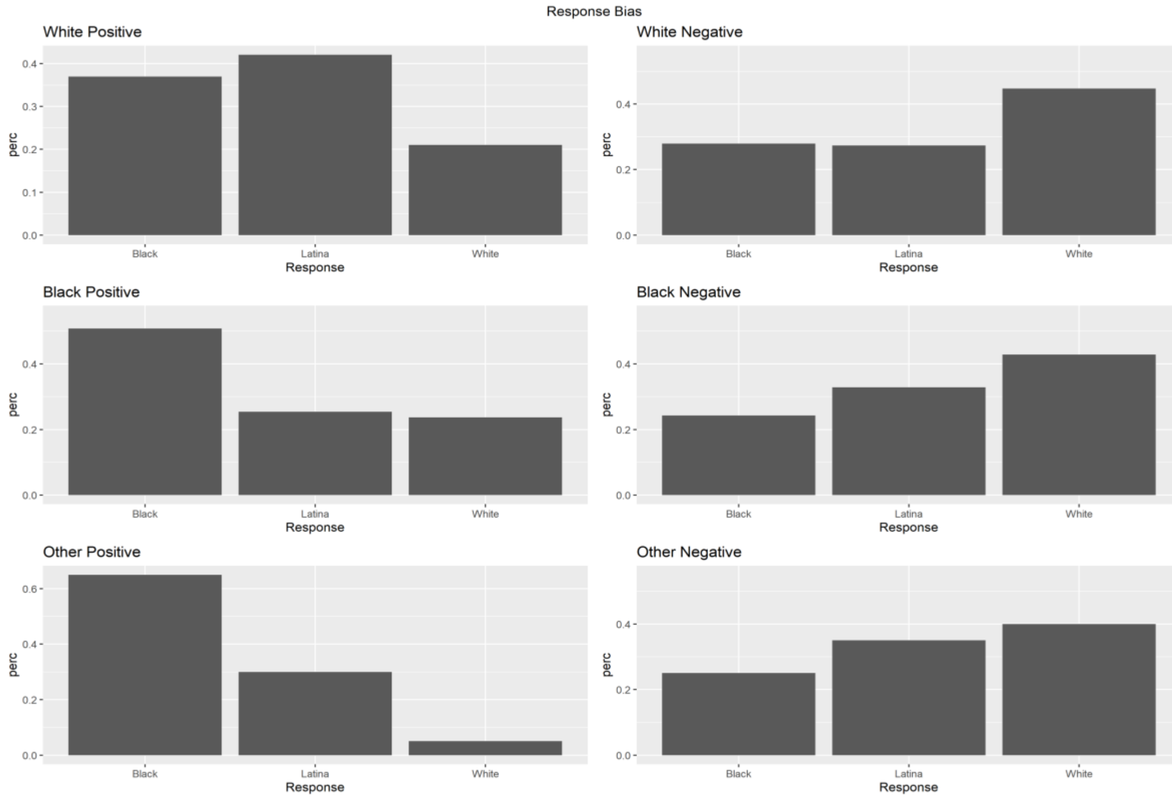

Figure 8: Survey Response Model Exp. 2: This figure shows participant responses to the survey. In the first row you see how white participants chose for positive questions v. negative questions. In the second row you see how Black participants chose for positive questions and negative questions. Finally, in the third row you see how other races chose for positive v. negative responses. It should be noted that participants were looking at the photos when they were responding.

Discussion

The accuracy results in Experiment 2 showed a different pattern from Experiment 1. In the Experiment 1, each group of participants did better with the group nearest to their own in-group identity. In Experiment 2, however, each group did better with the speaker opposite their identity. This could possibly be a result of confusion from the picture mismatch, but we cannot be sure. It is possible that with the White speaker and the White photo matching in the first experiment, and the other two options being minority groups, that possibly the match/mismatch was not as obvious but in this experiment with the White speaker being a part of the mismatch, it could have confused participants more.

The survey data revealed an increase in the Black picture being chosen for positive valence survey questions by all races and white being chosen for negative valence survey questions for all races. One uncontrolled aspect of this experiment that may be important to consider is the timing of when it was administered: directly following our summer of Black Lives Matter protests and education, which included the murder of George Floyd in Minneapolis, Breonna Taylor in Louisville, Ahmaud Arbery in Georgia and widespread protests and conversations that these violent acts triggered. This timing and the apparently opposite results obtained raise the concern that people were being especially conscious of the choices they were making when participating in this experiment.

Experiment 3

For experiment three, the methods and procedure are the same with the exception being that in this experiment, the Latina photo matches the audio with the Latina speaker, whilst the White and Black photos and audios are switched.

Results

Data Cleaning

Data were cleaned based on two separate measures, the first being accuracy and the second being survey responses. Participants with low mean accuracy scores (50% or less) on the comprehension questions were excluded, resulting in the removal of four participants. We then further removed participants who either put N/A for all their responses or selected the same answer for all nine survey questions. This removed thirteen participants leaving us with a total of sixty-one participants in Experiment Three.

Accuracy Analysis

For the comprehension question accuracy analysis, we followed the same methods used in experiment one. The model included fixed effects of Participant Race and Speaker Race as well as their interaction. These fixed effects were sum coded valence -.5 and .5 and then were treatment coded with Other being the baseline for both fixed effects as it is our control across experiments. The model is shown below.

Regression Equation: Accuracy ~ Participant Race * Speaker + (1|Participant) + (1|Item)

Audio Number	**Race of Woman Speaking**	**Picture Displayed**
R1	Latina	Latina
R2	Latina	Latina
R3	Latina	Latina
R4	Latina	Latina
R5	White	Black
R6	White	Black
R7	White	Black
R8	Black	White
R9	Black	White
R10	Black	White

Table 8: Audio/Visual Match/Mismatch for Exp. 3

Results from the accuracy analysis are reported in Figure 10 and Table 8. Black participants had higher accuracy scores with both the Black (est. = 0.545, z = 2.349 p = <.02) and White speaker (est. = 0.71, z = 3.123, p = < .002) in comparison to the Other group.

Additionally, White participants performed significantly better with the White speaker in comparison to the Other group (est. =.646, z = 5.119, p < .001).

	Estimate	Std. E	z value	p value	
White Participants	-0.003974	0.240446	-0.017	0.98681	
Black Participants	-0.182435	0.420915	-0.433	0.66471	
Black Speaker	-0.125443	0.406158	-0.309	0.75743	
White Speaker	-0.455597	0.405601	-1.123	0.26132	
White Participant * Black Speaker	0.054968	0.127327	0.432	0.66595	
Black Participant * Black Speaker	0.54572	0.232348	2.349	0.01884	*
White Participant * White Speaker	0.646652	0.126327	5.119	3.07E-07	***
Black Participant * White Speaker	0.709102	0.227075	3.123	0.00179	**

Table 9: Experiment 3 Accuracy Model

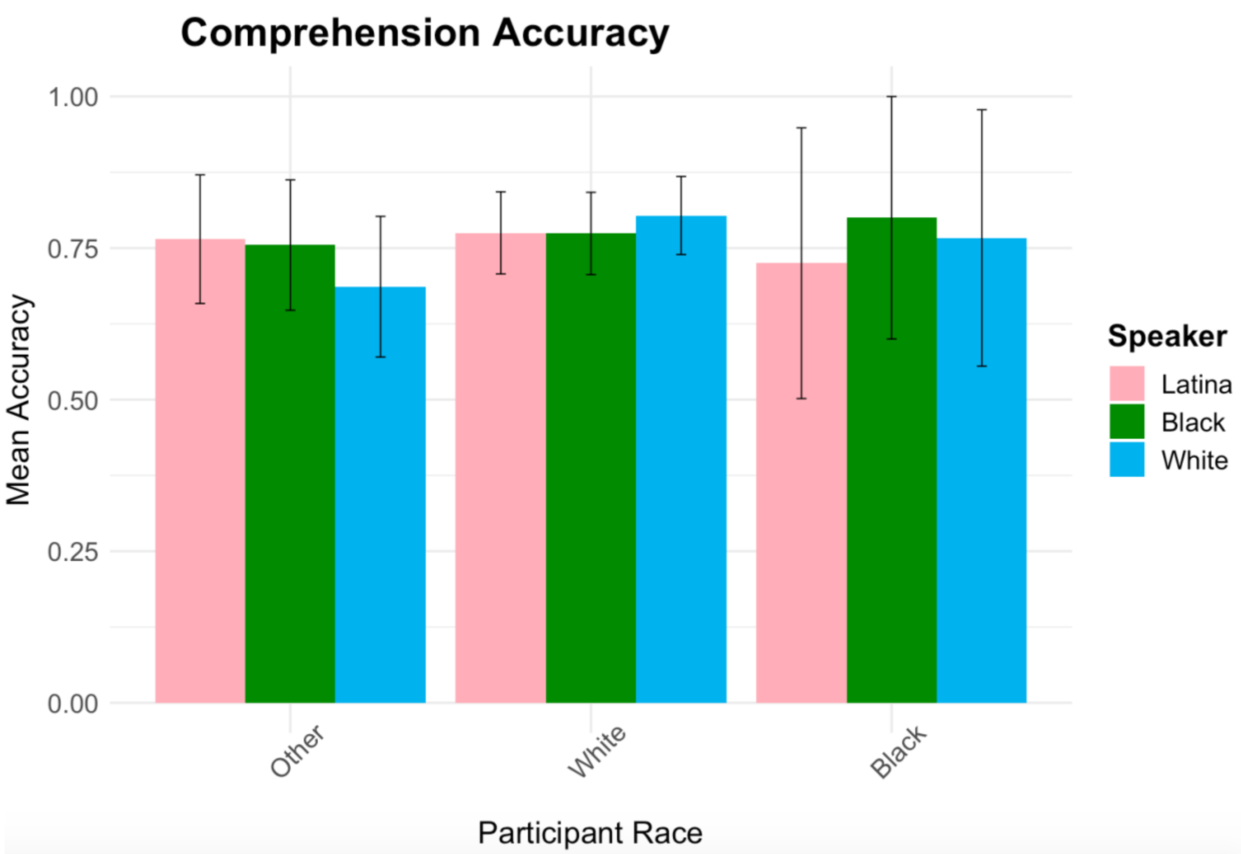

Figure 9:Accuracy Means Comparing Participant Race with Speaker Race

Survey Analysis

Similarly to experiment one and two, three separate regression models were run for each race. There was not a large enough Latina/o representation for their own model, so the "Other" category was created here.

The results displayed in Table 10 provide the results of the three regression models run in this experiment on the survey data. Each panel of Table 7 displays how each of the three groups responded in the survey. Figure 10 shows participant responses to the survey questions. The first row shows how white participants responded for positive valence questions vs. negative valence questions. The second row shows how Black participants responded for positive valence questions vs. negative valence questions. The third row shows how Other races responded for positive-valence questions vs. negative-valence questions. The x-axis on each panel represents the photos the participants were viewing and choosing from when they were responding to each question. The y-axis represents the percentage of total responses given by choosing the given photo. Percentages within each panel add up to 100%.

The results in Table 10 show that in the first regression model "White Response" (i.e., when participants chose the White photo for their answers), no main effects for valence were found, nor were any simple effects of participant race found or significant interactions between groups. However, in regression model two "Black Response" (i.e., when participants chose the Black photo for their answers) there is a main effect of valence for the response "Black" for positive survey questions overall (est. =0.374, t = 4.547, p = 6.99e-06) illustrated in the middle two panels labeled "Black Positive" and "Black Negative" of Figure 10. In regression model three "Other Response" (i.e., when participants chose the Latina photo for their answers) there was a main effect of valence for the response "Latina" for negative survey questions overall (est.

= -0.389, *t* = -4.536, *p* =7.34e-06) illustrated in the left three panels labeled “White Negative”, “Black Negative” and “Other Negative” of Figure 10. These results are summarized below.

White Response						
		Estimate	Std. Error	z value	Pr (>|t|)	
	Valence	0.05716	0.41054	0.139	0.889	
	Participant Race White	0.15977	0.23879	0.669	0.503	
	Participant Race Black	0.56603	0.44306	1.278	0.201	
	Valence: Participant Race White	0.14725	0.47757	0.308	0.758	
	Valence: Participant Race Black	0.12516	0.88612	0.141	0.888	
Black Response						
		Estimate	Std. Error	t value	Pr (>|t|)	
	Valence	0.37442	0.08234	4.547	6.99E-06	***
	Participant Race White	-0.03156	0.04831	-0.653	0.514	
	Participant Race Black	-0.05729	0.09536	-0.601	0.548	
	Valence: Participant Race White	-0.1453	0.09661	-1.504	0.133	
	Valence: Participant Race Black	-0.0997	0.19072	-0.523	0.601	
Other Response						
		Estimate	Std. Error	t value	Pr (>|t|)	
	Valence	-0.385945	0.085078	-4.536	7.34E-06	***
	Participant Race White	-0.002129	0.049911	-0.043	0.966	
	Participant Race Black	-0.069346	0.098531	-0.704	0.482	
	Valence: Participant Race White	0.112869	0.099823	1.131	0.259	
	Valence: Participant Race Black	0.067263	0.197062	0.341	0.733	

Table 10:: Experiment 3 Survey Model

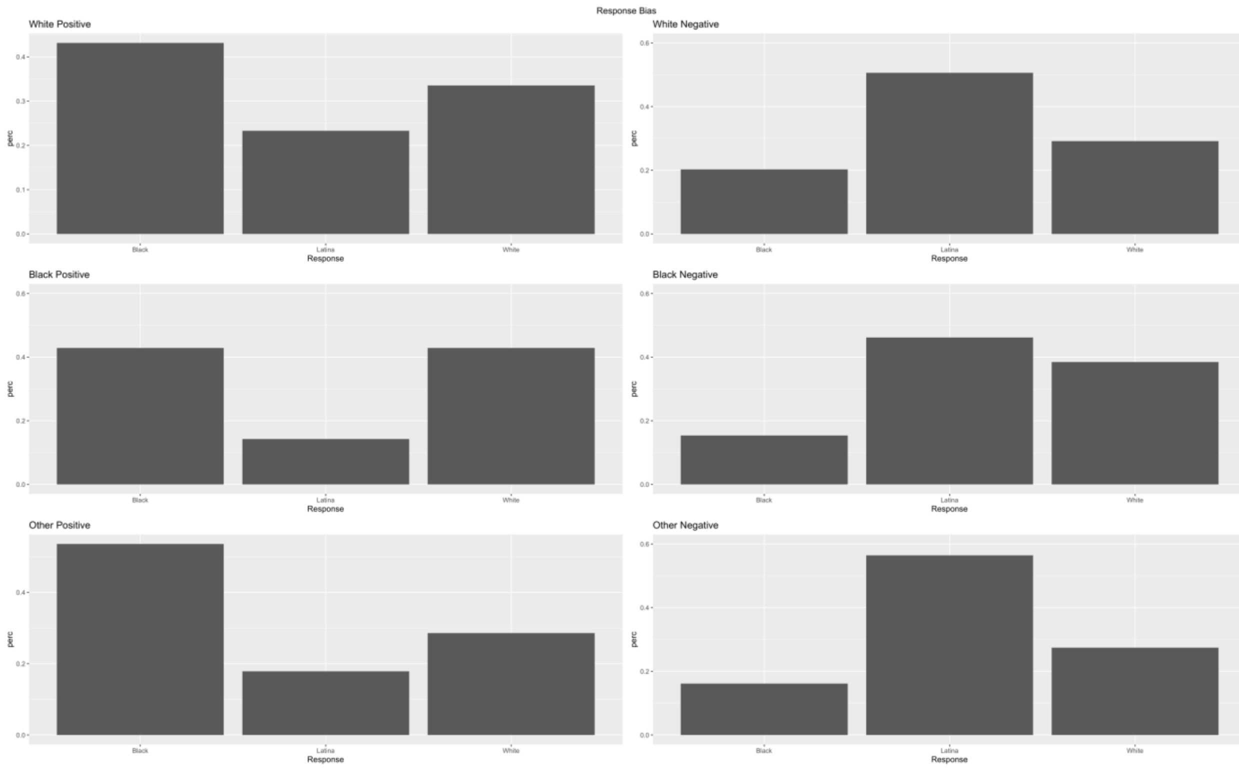

Figure 10: Survey Response Model Exp. 3: This figure shows participant responses to the survey. In the first row you see how white participants chose for positive questions v. negative questions. In the second row you see how Black participants chose for positive questions and negative questions. Finally, in the third row you see how other races chose for positive v. negative responses. It should be noted that participants were looking at the photos when they were responding.

Discussion

These results in Experiment 3 basically patterned similarly to Experiment 1. Looking at the accuracy model, it can again be seen that each group did better with the speaker nearest their identity, although Black participants did quite well with both the Black and White speaker. This pattern is interesting, considering that the White and Black voices were paired with the Black and White photos, respectively. These results raise further questions about the mismatching of audio and visual input, suggesting that *both* played a role in heightening attention or lowering affective barriers—or both—to comprehension.

As in Experiment 2, there was a large proportion of Black responses in the positive valence questions in in the surveys; however, it should be noted that the Black picture in this experiment was paired with the White speaker. Keep in mind that that the Black photo was paired with the White speaker in this experiment. As such, this result might also suggest that both photo and voice were contributing to the attitude results as well as comprehension accuracy results. The Latina picture and voice were matched in this experiment and received the most negative attributes from White and Black participants. On the other hand, the Other participants never chose the Latina for a negative valence question.

CHAPTER FOUR: GENERAL DISCUSSION

This study sought to identify the perceptibility of racial accents within an academic setting and to understand some of the potential effects that come with racial recognition linguistically. The results of both comprehension accuracy and attitude survey analyses strongly support the argument that accent bias is real and measurable.

The results of Experiments 1 and 3 most closely align with predictions of social identity theory (Tajfel & Turner, 1979). We found that participants were more accurate when listening to the speaker nearest their own identity, possibly implying that there was some sort of recognition taking place as they listened to the speakers despite the picture mismatches. This supports the findings in Bestelmeyer, Belin, and Ladd's (2015) study, which found that participants had a heightened sensitivity to in-group speakers when listening to audio recordings from different speakers. It appears that our picture mismatch manipulation did not dramatically alter this general pattern of results, at least in Experiment 1 or 3. Experiment 2 proved to be an anomaly however, with each group doing better with a race they did not identify with. As noted above, a possible reason for this is the timing of the experiment. It was run in early August, after a summer of racial tension and public calls for study and education about race and implicit biases that contribute to racism. If this is the reason for the aberrant results in Experiment 2, it would suggest that participants were focused on appearing non-biased, according to the Wang et al. (2012) study that found that suppressing one's biases requires both a cognitive and an affective effort on the listener's part. It should be noted, however, the participants identifying as Black had the lowest accuracy scores overall which could be interesting as they were arguably more deeply affected by the events of the summer. They could have been a different kind of bias taking place for these participants as they walked through the experiment. Although it is possible that these participants simply had unique perspectives surrounding race and language, I suspect that a conscious effort to not appear racially biased is more likely what happened in Experiment 2.

Furthermore, in general the non-white speakers were more likely to be chosen for negative valence survey questions. In Experiment 1, the majority of our negative question choices were for the Black speaker, and in Experiment 3 the Latina speaker received the majority of the negative question responses. Again, the only experiment where this pattern was not observed was in Experiment 2, where participants chose the White speaker for negative questions. These results support the results in studies by both Hosoda and Stone-Romero (2010) and Shah (2019), that showed that people are sensitive to and tend to react negatively to accents that differ from the expected, even if the speech is understood. It also could have been a reaction to the photos, but it is important that we do not separate one from the other. This is a clear example of racial bias, one that can be damaging within the realm of academia and beyond. These kinds of biases can lead to negative evaluations, negative interactions, and negative experiences overall within the community. This could result in a lack of retention of Black and Latina/o graduate students and faculty. According to the National Center for Education Statistics (2018) Black women hold only 2% of tenured professor positions and Latina women hold only 1%.

There are clear limitations to this work that could have influenced the results, such as the aforementioned racial tensions of the time. Race-based studies are often deeply rooted in context and the feelings of the time in which they were done. Another limitation is the demographic profile of the participants who participated in the experiments. In the context of academic evaluations, in all three experiments there were a large number of participants who identified as age 35 or older, and although this is not necessarily a bad thing, in the setting of academia, where many student evaluators are between 18-25, the results might not be reflective of the population most involved the in the evaluation process.

The present study suggests many different avenues for future work. Replicating this study with other non-white groups (such as Asian/Indian American, Indigenous Americans, etc.) is one such avenue, as the speakers could be an area of interest for expansion of the work. Moreover, a deeper analysis into the picture and voice impacts of the study could be done. One possibility would be re-running the experiments with just the voices or just the photos and seeing if results replicate. This study could be performed to examine accent bias in children as well, should a more developmental route be of interest.

To conclude, the main question in this study was whether racial accents are perceptible in quasi-academic contexts, and how much visual cues to a speaker's race influence how or whether they are heard and the content of their speech comprehended. Although the results were not completely consistent across experiments, and the effects of the picture/voice mismatch are still somewhat unclear, findings show that participants did perform with greater accuracy when listening to their "in-group" speaker. It is clear that racial accents are perceptible in quasi-academic contexts and consequently may work against certain speakers, as supported by the generally higher proportion of negative attributions to non-white speakers in the survey results.

REFERENCES

Andrews, E. (2019, January 29) 8 Reasons Why Rome Fell. https://www.history.com/news/8-reasons-why-rome-fell

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Baugh, J. (2018). Ethnolinguistic Assertions Regarding People Who Allegedly "Talk White," or "Talk Black". In B. Evans, E. Benson, & J. Stanford (Eds.), *Language Regard: Methods, Variation and Change* (pp. 183-196). Cambridge: Cambridge University Press. doi:10.1017/9781316678381.011

Baugh, J. (1999). Linguistic Profiling. In S. Makoni, G. Smitherman, A. F. Ball, & A. K. Spears (Eds.), Black linguistics: Language, society and politics in Africa and the Americas, 155-168. CITY, STATE: Routledge

Bestelmeyer, P., Belin P., Ladd, D. (2014) A Neural Marker for Social Bias Toward In-Group Accents. Cerebral Cortex.

Boersma, P., Weenink, D. (2021). Praat. Version 6.1.40, retrieved from http://www.praat.org/

Clark, K. B., & Clark, M. P. (1947). Racial identification and preference in Negro children. In T. M. Newcomb & E. L. Hartley (Eds.), Readings in social psychology (pp. 602– 611). New York, NY: Holt, Rinehart & Winston.

Conference on College Composition & Communication. (July, 2020) *This Ain't Another Statement! This is a DEMAND for Black Linguistic Justice!* https://cccc.ncte.org/cccc/demand-for-black-linguistic-justice?fbclid=IwAR06-zx04ep9fusuoCj9kwlBHwbcLM3FEY8YNfq7DeaC0Hs0otTIJbP3k84

Drummond A (2013) Ibex Farm. Available: http://spellout.net/ibexfarm/.

Gibbons, S. (2019, October 4) *Can medical care exist without plastic?* https://www.nationalgeographic.com/science/article/can-medical-care-exist-without-plastic

Grogger, J. (2009) Speech Patterns and Racial Wage Inequality. *The Journal of Human Resources.* Vol 46.

Hao, Y. (2018, February 7). *YGDP members speak about linguistic prejudice.* https://ling.yale.edu/news/ygdp-members-speak-about-linguistic-prejudice#:~:text=Linguistic%20prejudice%20is%20a%20form,%2C%20generations%2C%20and%20ethnic%20groups

Hosoda, M., Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology, 25*(2), 113–132. https://doi.org/10.1108/02683941011019339

.

King, S., Kinzler, K. (2020, July 14). *Op-Ed: Bias against African American English speakers is a pillar of systemic racism.* https://www.latimes.com/opinion/story/2020-07-14/african-american-english-racism-discrimination-speech

Klein, C. (2017, October 25) *Halloween Was Once So Dangerous That some Cities Considered Banning It.* https://www.history.com/news/halloween-was-once-so-dangerous-that-some-cities-considered-banning-it

Little, B. (2020, January 19) *How 22-Year-Old George Washington Inadvertently Sparked a World War*. https://www.history.com/news/george-washington-french-indian-war-jumonville

Marks, J. (2019, July 8) *How Did the American revolution Influence the French Revolution?* https://www.history.com/news/how-did-the-american-revolution-influence-the-french-revolution

Merriam-Webster. (n.d.). Standard English. In *Merriam-Webster.com dictionary*. Retrieved January 18, 2021, from https://www.merriam-webster.com/dictionary/Standard%20English

Pruitt, S. (2019, October 22) *How Are Socialism and Communism Different?* https://www.history.com/news/socialism-communism-differences

Pruitt, S. (2019, September 25) *How T-Rex Got Its Powerful Bite*. https://www.history.com/news/tyrannosaurus-rex-bite-discovery

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

Sanchez, C., Khan, S. (2016) Instructor accents in online education and their effect on learning and attitudes. *Journal of Computer Assisted Learning*. Vol 32, 494–502

Shah, A. (2019) Why are Certain Accents Judged the Way they are? Decoding Qualitative Patterns of Accent Bias. *Advances in Language and Literary Studies*. ISSN: 2203-4714

Smith, P. (2020, August, 18) *Why for Black speakers, despite what they are told, using 'Standard English' will not lead to acceptance. https://blogs.lse.ac.uk/usappblog/2020/08/18/why-for-black-speakers-despite-what-they-are-told-using-standard-english-will-not-lead-to-acceptance/#Author*

U.S. Department of Education, National Center for Education Statistics. (2020). *The Condition of Education 2020* (NCES 2020-144), Characteristics of Postsecondary Faculty.

Wei-Haas, M. (2019, October 28) *Controversial new study pinpoints where all modern humans arose*. https://www.nationalgeographic.com/science/article/controversial-study-pinpoints-birthplace-modern-humans

Wei-Haas, M. (2019, October 2) *Earth's magnetic field flips much more frequently than we thought. https://www.nationalgeographic.com/science/article/earths-magnetic-field-flipped-more-times-scientists-thought*

Wei-Haas, M. (2019, October 3) *New 'iron dragon' pterosaur found in Australia.* https://www.nationalgeographic.com/science/article/new-iron-dragon-pterosaur-found-australia

APPENDIX A: STIMULI

Racial Demographics

Experiment	# of Participants	Match	Mismatch	Black	White	Other
1	80 -> 67	WV+WP	BV+LP, LV+BP	20	42	18
2	80 -> 68	BV+BP	WV+LP, LV+WP	19	50	11
3	80 -> 61	LV+LP	BV+WP, WV+BP	9	50	21

Comprehension Texts and Questions

Yet for all the ways plastic has revolutionized the medical industry over the past century, it's now being scrutinized for what happens after it's done its job. Plastic can easily end up in marine environments, where it breaks down into tiny particles called microplastics that have yet-to-be-determined health consequences. And the fossil fuels required to produce those plastics can contaminate air and water. Increasingly, say medical care providers, the unfettered use of plastic is conflicting with a doctor's promise to do no harm, but in facilities awash in blood and pathogens, is avoiding plastic even possible? (15.4, ARI)

1. Can the fossil fuels produced by plastic pollute water? (Y/N)

2. Does plastic use conflict with a doctor's promise to do no harm? (Y/N)

Named *Ferrodraco lentoni*, the new fossil is far from a full skeleton; it includes parts of the upper and lower jaw, five partial neck bones, sections of both wings, and many teeth. Despite pterosaur finds in Australia being exceedingly rare, these fossils are exceptionally well preserved. Described today in *Scientific Reports*, the newly found pterosaur is roughly 96 million years old, based on previously reported ages of the rock formation thought to entomb the creature's bones. Its closest relatives, pterosaurs of the group Anhangueria (an-hang-GWEHR-ee-ah), are thought to have died out by 94 million years ago. Though its precise age remains uncertain, *Ferrodraco* joins a number of exciting new Australian discoveries from the age of dinosaurs, including the most complete dinosaur fossil preserved as opal. (15.7, ARI)

1. The recently discovered fossil includes just one of the creature's wings. (T/F)

2. One dinosaur fossil is completely preserved as emerald. (T/F)

A protective magnetic bubble shields Earth from radiation that's constantly streaming from the sun. In the planet's 4.6-billion-year history, the field has frequently flipped, swapping magnetic north and south, and some research suggests that another flip may be on the geological horizon. While fears of a looming geomagnetic apocalypse are overblown, a magnetic reversal could have many damaging impacts, from increased radiation exposure to technological disruptions, which makes understanding these historic flips more than just a scientific curiosity. The planet usually experiences 26 magnetic pole reversals every million years—more than five times the rate seen in the last 10 million years. (14.9, ARI)

1. The planet has experienced 26 magnetic pole reversals every two million years. (T/F)

2. Could a magnetic flip increase radiation exposure and create technological disruptions? (Y/N)

In his 1875 writing, *Critique of the Gotha Program*, Karl Marx summarized the communist philosophy in this way: “From each according to his ability, to each according to his needs.” By contrast, socialism is based on the idea that people will be compensated based on their level of individual contribution to the economy. Unlike in communism, a socialist economic system rewards individual effort and innovation. Social democracy, the most common form of modern socialism, focuses on achieving social reforms and redistribution of wealth through democratic processes, and can co-exist alongside a free-market capitalist economy. (15.6, ARI)

3. Socialism is based on the idea that people will be compensated based on their level of individual contribution to the economy. (Y/N)

4. Social democracy cannot co-exist alongside a free-market capitalist economy. (T/F)

Technically, the “Jumonville Affair” was a military victory for Washington—but a diplomatic loss. The fact that he had attacked France, a country with which Britain was not at war, gave France a huge propaganda advantage. It also angered Joseph Coulon de Jumonville’s half-brother, a French military leader named Louis Coulon de Villiers, who, just over a month after his brother was killed, helped lead an attack on Washington’s Virginia Regiment at Fort Necessity. Unlike the Jumonville affair, the Battle of Fort Necessity was a military and diplomatic disaster for Washington. On July 3, a mix of French, Huron, Odawa and Iroquois fighters overwhelmed Washington’s men at their recently built fort. The Virginia Regiment, unable to drum up its own corps of native allies, was outnumbered and underprotected behind the small, flimsy Fort Necessity, which looked like a tall, circular fence and was situated in an open field. Ultimately, Washington surrendered to terms that included—unbenownst to him, because of a poor French translation—taking responsibility for the assassination of Jumonville. (15.4, ARI)

1. The Battle of Fort Necessity was a military and diplomatic success for Washington. (T/F)
2. Did Washington assassinate Jumonville? (Y/N)

Though most historians agree that the American Revolution impacted the French Revolution, which lasted from 1789-1799, some scholars debate the significance and extent of this effect. France, a country on the verge of financial collapse with an outdated feudal system and a wildly unpopular monarchy, was a powder keg waiting to explode, with or without the American war to serve as an example. Other political, social and religious factors also activated the French people’s appetite for change. Though there were clear differences between the motives for each revolt and how the two wars were fought, most experts believe that the war in America at least partly paved the way for France’s uprising. The Americans provided a working model of revolutionary success that wasn’t lost on the French. (15.1, ARI)

1. The French Revolution lasted from 1789-1800. (Y/N)
2. Did the Americans provide a working model of revolutionary success? (Y/N)

Researchers have concluded that T. Rex’s joints were likely tightly fused, making its skull extremely rigid and inflexible, able to withstand the tremendous force it used to bite down on its prey. This kind of skull resembles those of other powerful biters such as tigers and hyenas, as well as alligators and crocodiles, the biggest of which have a bite force of around 3,700 pounds, the strongest that has been measured among living animals. Scientists believe the most powerful bite force ever belonged to now-extinct giant crocodiles, which measured 35 to 40 feet long and may have had a bite force of up to 18,000 pounds. In addition to shedding new light on the anatomy of one of the prehistoric world's apex predators, the new study may impact modern medicine (both animal and human) by providing a better understanding of how joints and ligaments work. (15.2, ARI)

1. T. Rex's skull was likely extremely rigid and inflexible. (Y/N)
2. Giant crocodiles measured up to 40 feet long. (T/F)

The Great Depression exacerbated Halloween mayhem, with mischief often devolving into vandalism, physical assaults and sporadic acts of violence. One theory suggests that excessive pranks on Halloween led to the widespread adoption of an organized, community-based trick-or-treating tradition in the 1930s. This trend was abruptly curtailed, however, with the outbreak of World War II, when sugar rationing meant there were few treats to hand out. At the height of the postwar baby boom, trick-or-treating reclaimed its place among other Halloween customs. It quickly became standard practice for millions of children in America's cities and newly built suburbs. No longer constrained by sugar rationing, candy companies capitalized on the lucrative ritual, launching national advertising campaigns specifically aimed at Halloween. (15.2, ARI)

1. Excessive pranks on Halloween may have led to the tradition of trick or treating in the 1920's. (T/F)
2. Candy companies were once constrained by sugar rationing due to the Great Depression. (T/F)

Set in the middle of a harsh desert in southern Africa, a lush landscape would have been an appealing place for early humans to call home. Now, a controversial new study in *Nature* argues that an oasis of this kind, known as the Makgadikgadi–Okavango wetland, was not just any home, but the ancestral "homeland" for all modern humans today. The study revives a long-simmering debate about exactly where in Africa modern humans emerged, and it has drawn sharp criticism from several scientists. They point out that although all humans alive today have mitochondrial DNA passed on from a common ancestor—a so-called Mitochondrial Eve—this is just a tiny fraction of our total genetic material. So even if the proposed founder population described in the new study is the source of our mitochondrial DNA, many others likely contributed to today's genetic pool. (15.9, ARI)

1. Could this wetland have been the ancestral homeland for all modern humans? (Y/N)
2. Mitochondrial DNA represents the largest amount of our DNA. (T/F)

The fate of Western Rome was partially sealed in the late third century, when the Emperor Diocletian divided the Empire into two halves—the Western Empire seated in the city of Milan, and the Eastern Empire in Byzantium, later known as Constantinople.

The division made the empire more easily governable in the short term, but over time the two halves drifted apart. East and West failed to adequately work together to combat outside threats, and the two often squabbled over resources and military aid. As the gulf widened, the largely Greek-speaking Eastern Empire grew in wealth while the Latin-speaking West descended into economic crisis. Most importantly, the strength of the Eastern Empire served to divert Barbarian invasions to the West. Emperors like Constantine ensured that the city of Constantinople was fortified and well guarded, but Italy and the city of Rome—which only had symbolic value for many in the East—were left vulnerable. (15.3, ARI)

1. The Western Empire was seated in the city of Byzantium. (T/F)
2. Did the East and West argue over resources and military aid? (Y/N)

READABILITY MEAN:15.37

READABILITY STANDARD DEVIATION: .3

WC MEAN: 127

WC SD: 25.6

Survey Questions

- Which speaker was most comprehensible?
- Which speaker was least comprehensible?
- Which speaker had the strongest accent?

- Which speaker seemed more educated?
- Which speaker seemed more trustworthy?
- Which speaker seemed least trustworthy?
- Which speaker seemed least educated?
- Which speaker seemed most likeable?
- Which speaker seemed least likeable?

APPENDIX B: R CODE

Experiment 1

```
#read in the accuracy data
acc_data <- read.csv("Acc_Data.csv")

#check data types of columns
str(acc_data) #looks good, we need accuracy to be numeric or integer

#from visual inspections, we need to delete the following participants:
bb9ff46509d92f11c5eb8dd4f7b41f89, d9b0919521aed5a4bdc74a6b7d16191f,
e9ac1158e9eca23773bf2a50b42bca71, fc525fbab2510c99516da820f92a1178,
423d34cd435b30b215451e5c0cac27f1 for bad survey answers

bad_survey <- c("bb9ff46509d92f11c5eb8dd4f7b41f89",
"d9b0919521aed5a4bdc74a6b7d16191f", "e9ac1158e9eca23773bf2a50b42bca71",
"fc525fbab2510c99516da820f92a1178", "423d34cd435b30b215451e5c0cac27f1")

acc_data$bad_survey <- acc_data$Participant %in% bad_survey


library(plyr)
acc_data <- ddply(acc_data[acc_data$bad_survey=="FALSE",], .())
#check
unique(acc_data$Participant)
#81 - 5 = 76, so we're good

#check overall accuracy mean
mean(acc_data$Accuracy)

#mean accuracy is .656...that's pretty bad - some of these questions were difficult and some
people probably didn't really try to answer...I noticed some pretty low response times to these
questions for example, let's start there

#check mean response times by participant
bypart_rtmeans <- ddply(acc_data, .(Participant), summarize, mean=mean(RT), sd=sd(RT))
range(bypart_rtmeans$mean)

#check accuracy for item, maybe one question in particular was very difficult
byitem_accmeans <- ddply(acc_data, .(Item), summarize, mean=mean(Accuracy),
sd=sd(Accuracy))

#there were some tricky items that had less than 50% accuracy...if it's below chance, I wouldn't
include it when making my accuracy cutoffs, so let's remove R2 & R5 from the dataset

bad_items <- c("R2", "R5")
```

acc_data$item_cleaning <- acc_data$Item %in% bad_items #this creates a boolean T/F row of whether the item is good or bad

#now remove the bad items
itemcleaned_acc <- ddply(acc_data[acc_data$item_cleaning=="FALSE",], .())

#now let's check overall accuracy again
mean(itemcleaned_acc$Accuracy) #70% accuracy, that's better! now let's see the sd
sd(itemcleaned_acc$Accuracy) #.45 ... with accuracy, you usually don't clean by an sd criterion for this reason, we'd lose all of our data if we did that

#by participant means
bypart_accmeans <- ddply(itemcleaned_acc, .(Participant), summarize, mean=mean(Accuracy), sd=sd(Accuracy))

I think we can safely remove any participants with accuracy less than or equal to 50%...that's chance level, so we're going to lose some data here

bad_participants <- ddply(bypart_accmeans[bypart_accmeans$mean<=.5,], .())
#with this cutoff we lose 8 participants from accuracy...if you want to change it, just change the number above to whatever your cutoff is
```

**Now we want the survey data only with participants with high enough accuracy, then clean for sketchy survey responses**
```{r}
#first, read in the data
survey_data <- read.csv("Survey_Data.csv")

#now, let's clean out those participants

survey_data$badacc <- survey_data$Participant %in% bad_participants$Participant

survey_acccleaned <- ddply(survey_data[survey_data$badacc=="FALSE",], .())
#double check our code by checking we have 81 - 8 = 73 participants
unique(survey_acccleaned$Participant) #73, so we're good

#and now let's clean out those original 5 that gave bad answers to survey question
survey_acccleaned$bad_survey <- survey_acccleaned$Participant %in% bad_survey
survey_partcleaned <- ddply(survey_acccleaned[survey_acccleaned$bad_survey=="FALSE",], .())
#check
unique(survey_partcleaned$Participant)
#73-5 = 68 so we're good

```
#Kiel says to delete RTs greater than 2.5 sds above mean by participant, so now let's delete just those observations (we're done removing whole participants from the data)

#to do this, we need to arrange your data differently so that the reaction times and responses are side by side...this is complicated but doable
library(tidyverse)
rt_rows <- ddply(survey_partcleaned[survey_partcleaned$Sentence=="_REACTION_TIME_",], .())
survey_partcleaned$arrange_bool <- ifelse(survey_partcleaned$Sentence=="_REACTION_TIME_",1,0)
answer_rows <- ddply(survey_partcleaned[survey_partcleaned$arrange_bool=="0",], .())
answer_rows <- answer_rows %>% select(Item, Response)
colnames(rt_rows)[colnames(rt_rows)=="Response"] <- "RT"
colnames(rt_rows)[colnames(rt_rows)=="Item"] <- "RT_Item"
survey_arranged <- cbind(rt_rows,answer_rows)

#get rid of R10 item and R3
R10 <- c("Survey10","Survey3")
survey_arranged$item_delete <- survey_arranged$Item %in% R10
survey_arranged <- ddply(survey_arranged[survey_arranged$item_delete=="FALSE",], .())
unique(survey_arranged$Item)

#now we're ready to go, first get by participant RT means
#make RT column numeric
survey_arranged$RT <- as.numeric(survey_arranged$RT)
RT_means <- ddply(survey_arranged, .(Participant), summarize, mean=mean(RT), sd=sd(RT))
RT_means$lowcutoff <- RT_means$mean + 2.5*(RT_means$sd)
RT_means$highcutoff <- RT_means$mean - 2.5*(RT_means$sd)
survey_RT <- merge(survey_arranged,RT_means)
survey_RT <- ddply(survey_RT[survey_RT$RT<=survey_RT$lowcutoff,], .())
clean_survey <- ddply(survey_RT[survey_RT$RT>=survey_RT$highcutoff,], .())

#now let's see how much we lost
dim(survey_RT)-dim(clean_survey)
#lol, we lost 1 data point, so less than 1% lost from RT cleaning
```

Descriptives and Preparing for Inferential Stats

We want to see descriptives for how people in general responded to each speaker

```{r}
#first, we need a list of which items were "negative" and which items were "positive"...we can look more directly at what the items were asking later. The strategy here is to make a new binary variable consisting of "bad" and "good", and then we can take the mean ratings that way...we can also break it down into trustworthy, etc. if you want, this is completely up to you
```

```
pos_item <- c("Survey1", "Survey4", "Survey5", "Survey8")

clean_survey$Valence_bool <- clean_survey$Item %in% pos_item

clean_survey$Valence <- ifelse(clean_survey$Valence_bool=="TRUE","Positive","Negative")

#what about race of participant?
unique(clean_survey$Race)
#participant race dataframes just in case
white_data <- ddply(clean_survey[clean_survey$Race=="White",], .())
black_data <- ddply(clean_survey[clean_survey$Race=="Black",], .())
hispanic_data <- ddply(clean_survey[clean_survey$Race=="Hispanic",], .())

#exclude Other participant races
main_races <- c("White", "Hispanic", "Black")
clean_survey$Race_other_check <- clean_survey$Race %in% main_races
clean_survey$Race_Group <- ifelse(clean_survey$Race_other_check=="FALSE",
"Other","NA")
other_data <- ddply(clean_survey[clean_survey$Race_Group=="Other",], .())

#make other participant race dataframe
other_data <- other_data %>% select(-Race_other_check,-Race_Group)

white_data$Participant_Race <- "White"
black_data$Participant_Race <- "Black"
hispanic_data$Participant_Race <- "Hispanic"
other_data$Participant_Race <- "Other"

clean_survey_grprace <- rbind(white_data,black_data,hispanic_data,other_data)

#We have an insane amount of columns, so let's clean that up
clean_survey_refresh <- clean_survey_grprace %>% select(Participant, Age, Gender, Race,
Participant_Race, Item, RT, Response, Valence)

count_(clean_survey_refresh, 'Item') #we're good, no more weird problems

unique(clean_survey_refresh$Gender)
#since genders just male and female in these data, no need to code further

#valence: positive = 1, negative = 0
clean_survey_refresh$Valence <- revalue(clean_survey_refresh$Valence,
c("Positive"="1","Negative"="0"))
clean_survey_refresh$Valence <- as.numeric(clean_survey_refresh$Valence)

#make dummy responses
```

clean_survey_refresh$white <- ifelse(clean_survey_refresh$Response=="White",1,0)
clean_survey_refresh$black <- ifelse(clean_survey_refresh$Response=="Black",1,0)
clean_survey_refresh$latina <- ifelse(clean_survey_refresh$Response=="Latina",1,0)

#remove NAs
clean_survey_refresh <- clean_survey_refresh[!(is.na(clean_survey_refresh$Response) | clean_survey_refresh$Response==""),]

#How many people per race participated?
##library(dplyr)
##count(clean_survey_refresh$Participant_Race)

###LOOK AT THIS DATAFRAME IF YOU WANT TO MAKE DESCRIPTIVE MEAN/SD TABLES
bias <- ddply(clean_survey_refresh, .(Participant_Race, Valence), summarize, meanwhite=mean(white), meanblack=mean(black), meanlatina=mean(latina), sdwhite=sd(white), sdblack=sd(black), sdlatina=sd(latina))

bias$Valence <- as.factor(bias$Valence)
bias$Valence <- revalue(bias$Valence, c("1"="Positive", "0"="Negative"))

white_bias <- ddply(bias[bias$Participant_Race=="White",], .())
black_bias <- ddply(bias[bias$Participant_Race=="Black",], .())
latina_bias <- ddply(bias[bias$Participant_Race=="Hispanic",], .())

latina_bias$Participant_Race <- revalue(latina_bias$Participant_Race, c("Hispanic"="Latina"))

head(clean_survey_refresh,10)
```
**Look at final dataframe**

We need columns for: Participant, Age, Part_Gender, Part_Race, Item, RT, Response, Valence, and 3 dummy variables for responses

### Response Graphic -- Follow this exactly for other experiments, descriptive purposes
```{r}

library(gridExtra)

unique(clean_survey$Race)
kiel_graph <- clean_survey

kiel_graph$Race <- revalue(kiel_graph$Race, c("Native American"="Other", "Hispanic"="Other", "Asian"="Other", "Black_White"="Other","Hispanic_White"="Other"))

```
unique(kiel_graph$Race)
#remove nas
kiel_graph <- kiel_graph[!(is.na(kiel_graph$Response) | kiel_graph$Response==""), ]


#Positive graphs
kiel_white <- ddply(kiel_graph[kiel_graph$Race=="White",], .())
kiel_white_pos <- ddply(kiel_white[kiel_white$Valence=="Positive",], .())
kiel_white_neg <- ddply(kiel_white[kiel_white$Valence=="Negative",], .())

kiel_white_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_white_pos)) -> kiel_white_pos_perc

kiel_white_pos_graph <- ggplot(kiel_white_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("White Positive")+
  ylim(0,.55)

kiel_white_pos_graph

kiel_black <- ddply(kiel_graph[kiel_graph$Race=="Black",], .())
kiel_black_pos <- ddply(kiel_black[kiel_black$Valence=="Positive",], .())
kiel_black_neg <- ddply(kiel_black[kiel_black$Valence=="Negative",], .())

kiel_black_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_black_pos)) -> kiel_black_pos_perc

kiel_black_pos_graph <- ggplot(kiel_black_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Black Positive")+
  ylim(0,.55)

kiel_black_pos_graph

kiel_other <- ddply(kiel_graph[kiel_graph$Race=="Other",], .())
kiel_other_pos <- ddply(kiel_other[kiel_other$Valence=="Positive",], .())
kiel_other_neg <- ddply(kiel_other[kiel_other$Valence=="Negative",], .())

kiel_other_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_other_pos)) -> kiel_other_pos_perc

kiel_other_pos_graph <- ggplot(kiel_other_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
```

```
  ggtitle("Other Positive")+
  ylim(0,.55)

kiel_other_pos_graph

#Negative Graphs
kiel_white_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_white_neg)) -> kiel_white_neg_perc

kiel_white_neg_graph <- ggplot(kiel_white_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("White Negative")+
  ylim(0,.55)

kiel_white_neg_graph

kiel_black_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_black_neg)) -> kiel_black_neg_perc

kiel_black_neg_graph <- ggplot(kiel_black_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Black Negative")+
  ylim(0,.55)

kiel_black_neg_graph

kiel_other_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_other_neg)) -> kiel_other_neg_perc

kiel_other_neg_graph <- ggplot(kiel_other_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Other Negative")+
  ylim(0,.55)

kiel_other_neg_graph

#combine graphs
grid.arrange(kiel_white_pos_graph,kiel_white_neg_graph,kiel_black_pos_graph,kiel_black_neg
_graph,kiel_other_pos_graph,kiel_other_neg_graph,nrow=3,top="Response Bias")
```

Response Analyses

Conduct separate analyses on each response

Fixed Effects: Age (2 levels, sum contrasts) interaction, Gender (treatment coded male baseline), & Participant_Race*valence (Other baseline, sum contrasts)

```{r}
analysis_data <- clean_survey_refresh

analysis_data$Age <- revalue(analysis_data$Age, c("35_Plus"="Old", "18-35"="Young"))
analysis_data$Age <- as.factor(analysis_data$Age)
contrasts(analysis_data$Age) <- c(.5, -.5) #young is coded as negative

analysis_data$Gender <- as.factor(analysis_data$Gender)

analysis_data$Participant_Race <- as.factor(analysis_data$Participant_Race)
analysis_data$Participant_Race <- factor(analysis_data$Participant_Race,
c("Other","White","Black","Hispanic"))

analysis_data$Valence <- as.factor(analysis_data$Valence)
analysis_data$Valence <- revalue(analysis_data$Valence, c("1"="Positive", "0"="Negative"))
contrasts(analysis_data$Valence) <- c(-.5, .5) #negative coded as negative

##Analysis

#White Response analysis

library(lme4)
library(lmerTest)

white_response_model1 <- glmer(white~Age*Gender*Valence*Participant_Race +
(1|Participant) + (1|Item), family="binomial", analysis_data)

#doesn't fit

summary(white_response_model1)

#remove Age from model

white_response_model2 <- glmer(white~Gender*Valence*Participant_Race + (1|Participant) +
(1|Item), family="binomial", analysis_data)

#singular fit - remove Gender
#not enough data to run such a complex model yet...maybe in the overall analysis

white_response_model3 <- glmer(white~Valence*Participant_Race + (1|Participant),
family="binomial", analysis_data)
```

```
#singular fit

summary(white_response_model3)

summary(analysis_data$Participant_Race)

#there are not enough hispanic participants...let's move those to other in this case...

analysis_data$Participant_Race <- revalue(analysis_data$Participant_Race,
c("Hispanic"="Other"))

new_white_response_model1 <- glmer(white~Age*Gender*Valence*Participant_Race +
(1|Participant) + (1|Item), family="binomial", analysis_data)

#drop age again

new_white_response_model2 <- glmer(white~Gender*Valence*Participant_Race +
(1|Participant) + (1|Item), family="binomial", analysis_data)

#singular fit, drop gender
str(analysis_data)
new_white_response_model3 <- glmer(white~Valence*Participant_Race + (1|Participant),
family="binomial", analysis_data)

summary(new_white_response_model3)

#mixed effects models aren't working, for some reason the random effects explain almost no
variance at all

white_response_reg_model1 <- lm(white~Age*Gender*Valence*Participant_Race,
data=analysis_data)

summary(white_response_reg_model1)

#some NAs, we should remove Age and Gender

white_response_reg_model2 <- lm(white~Valence*Participant_Race, data=analysis_data)

summary(white_response_reg_model2)

#white model --> marginally significant valence: positive items trend towards white response
#white model --> black participants rate nonwhite participants more in general


###Black Model
```

#we can skip right to regression

black_response_reg_model1 <- lm(black~Valence*Participant_Race, data=analysis_data)

summary(black_response_reg_model1)

#black model --> interaction of valence & white participants

interaction.plot(analysis_data$Valence, analysis_data$Participant_Race, analysis_data$black)

#black model --> white participants more likely to select black response for negative questions!!

###Latina model

latina_response_reg_model1 <- lm(latina~Valence*Participant_Race, data=analysis_data)

summary(latina_response_reg_model1)

#latina model --> trending that white participants choose them more often in general
#latina model --> black participants choose them more (may be interesting, since that waas black voice in this experiment)
#latina model --> interactions for both white and black with valence

interaction.plot(analysis_data$Valence, analysis_data$Participant_Race, analysis_data$latina)

#latina model --> white & black participants choose latina for positive items more, opposite trend for other participants
```

### Response Analysis Summary

**White participants choose black for negative items, and there seems to be a black & white trend toward selecting Latina for positive items**

### Accuracy
```{r}
acc_analysis <- read.csv("C:/Users/jkdem/Box Sync/School/Projects/ABC/ABC 1/Analysis/Acc_Data.csv")

acc_analysis$Speaker <- acc_analysis$Item

acc_analysis$Speaker <- revalue(acc_analysis$Speaker, c("R1"="White", "R2"="White", "R3"="White", "R4"="White", "R5"="Black", "R6"="Black", "R7"="Black", "R8"="Latina", "R9"="Latina", "R10"="Latina"))
unique(acc_analysis$Speaker)

part_race <- analysis_data %>% select(Participant,Participant_Race)

acc_merged <- merge(part_race,acc_analysis)
#First do accuracy
unique(acc_merged$Participant_Race)
#join hispanic and other
acc_merged$Participant_Race <- revalue(acc_merged$Participant_Race, c("Hispanic"="Other"))
#dummy code participant race
acc_merged$Black_part <- ifelse(acc_merged$Participant_Race=="Black",1,0)
acc_merged$Other_part <- ifelse(acc_merged$Participant_Race=="Other",1,0)
#speaker
unique(acc_merged$Speaker)

mean(acc_merged$Accuracy)

large_acc_model <- glmer(Accuracy~Participant_Race*Speaker+(1|Participant)+(1|Item),
family="binomial", acc_merged)

summary(large_acc_model)

#interaction plots

interaction.plot(acc_merged$Speaker, acc_merged$Participant_Race, acc_merged$Accuracy)
```

Accuracy Summary

White & Black participants were more accurate than other participants. Black participants had much higher accuracy for black speakers (Latina picture), and higher acc for latina compared with white speakers. White participants also had higher accuracy for white speakers. "Other" participants had higher accuracy for Latina speakers. So the general trend seems to be that everyone does better listening to their own race. This is assuming people aren't tricked by the picture.

Accuracy Descriptives
```{r}
#make the means dataframe
accuracy_means <- ddply(acc_merged, .(Participant_Race,Speaker), summarize,
mean=mean(Accuracy),sd=sd(Accuracy),nd = n_distinct(Participant))

library(ggthemes)

is.numeric(accuracy_means$mean)

acc_means_plot <- ggplot(accuracy_means,aes(y=mean,x=Participant_Race,bg=Speaker))+

```
  geom_bar(position="dodge",stat="identity")+
  geom_errorbar(aes(ymin=mean-sqrt((mean*(1-mean))/(nd-1)),ymax=mean+sqrt((mean*(1-mean))/(nd-1)), width=.1), position=position_dodge(width=.9),size=.3)+
  ggtitle("Accuracy Means")+
  ylab("Mean Accuracy")+
  xlab("Participant Race")+
  theme_calc()

.66+sqrt((.66*(1-.66))/22)

acc_means_plot

acc_means_plot + coord_cartesian(ylim=c(.5,.81))

```

**Experiment 2**

**Read in the Data**

```{r}
raw_survey <- read.csv("ABC2_Survey.csv")
raw_accuracy <- read.csv("ABC2_Accuracy.csv")

#need to remove 1eb8c57f3bef4272c20a629297961432 & 35ead1ca93afe3b07393553426cd66ec
#these two only answered one non-NA

bad_survey <- c("35ead1ca93afe3b07393553426cd66ec",
"1eb8c57f3bef4272c20a629297961432")

raw_accuracy$bad <- raw_accuracy$Participant %in% bad_survey

library(plyr)

#now we delete the bad people

clean_accuracy <- ddply(raw_accuracy[raw_accuracy$bad==FALSE,], .())

#let's look at the data format for each column
str(clean_accuracy)

mean(clean_accuracy$Accuracy) #61.86%

#means by participant to look for outlier participants

bypart_accmeans <- ddply(clean_accuracy, .(Participant), summarize, mean=mean(Accuracy))
hist(bypart_accmeans$mean)

#oof, that is a lot of bad participants, so let's see if there are certain problematic items
```

byitem_accmeans <- ddply(clean_accuracy, .(Items), summarize, mean=mean(Accuracy))
byitem_accmeans

#let's remove the items with lower than 50% accuracy and then re-inspect participant means

bad_item <- c("R5", "R2")

clean_accuracy$bad_item <- clean_accuracy$Items %in% bad_item

item_cleaned_acc <- ddply(clean_accuracy[clean_accuracy$bad_item==FALSE,], .())

bypart_accmeans_itemcleaned <- ddply(item_cleaned_acc, .(Participant), summarize,
mean=mean(Accuracy))

mean(item_cleaned_acc$Accuracy)

#following cutoff of 50% accuracy from experiment 1, remove bad participants

#create a list of participants for removal
remove_from_survey <- c("35ead1ca93afe3b07393553426cd66ec",
"1eb8c57f3bef4272c20a629297961432", "4a88f91331232acd5bee28d2543c83e6",
"75193f139818db5b0bb821dcf2bb3dd3", "eb6553d503add6cc9331b00f25c05301",
"34b0d0ec4879a8628b3e07be942028ad", "61eda04aa8160ee8555777c0f2350f11",
"74853a2f693ca05223845cb014607794", "8211ab199fc39a6bdc6dfc4352fcccbc",
"265c1fdf91d6ed7edda9e9c5c36c7e1f", "affebff3c0d7aabb7cc2a1551f009d25",
"f5155953a4c330762e9887d01a225b2d")

library(tidyverse)

```

**Wrangle the Survey Data**
```{r}
raw_survey$bad <- raw_survey$Participant %in% remove_from_survey

clean_survey <- ddply(raw_survey[raw_survey$bad==FALSE,], .())

unique(clean_survey$Participant) #67 participants

#look at survey data types
str(clean_survey)

#don't clean based on reaction times right now...we could do that later if needed, but only one
point was deleted last time, likely not an issue

```
pos_item <- c("Survey1","Survey4","Survey5","Survey8")

clean_survey$Valence_bool <- clean_survey$Items %in% pos_item
clean_survey$Valence <- ifelse(clean_survey$Valence_bool=="TRUE",1,0)

#participant race
str(clean_survey)
unique(clean_survey$Race)

#Black_White --> count as Black

clean_survey$Race <- revalue(clean_survey$Race, c("Black_White"="Other", "Native
American"="Other", "N/A"="Other", "Asian"="Other"))

unique(clean_survey$Race)

#now make Other the baseline for participant race (do this for all 3 experiments)

clean_survey$Participant_Race <- as.factor(clean_survey$Race)

summary(clean_survey$Participant_Race)

#we'll need to group hispanic with other again because there aren't enough...maybe in overall
anaysis

clean_survey$Participant_Race <- revalue(clean_survey$Participant_Race,
c("Hispanic"="Other"))

clean_survey$Participant_Race <- factor(clean_survey$Participant_Race,
c("Other","White","Black"))

#make response dummy variables

summary(as.factor(clean_survey$Response))

clean_survey$white <- ifelse(clean_survey$Response=="White",1,0)
clean_survey$black <- ifelse(clean_survey$Response=="Black",1,0)
clean_survey$latina <- ifelse(clean_survey$Response=="Latina", 1, 0)

#now delete unnecessary columns

analysis_data <- clean_survey %>% select(Participant, Items, Age, Gender, Participant_Race,
Valence, white, black, latina)

```

We need columns for: Participant, Age, Part_Gender, Part_Race, Item, Response, Valence, and 3 dummy variables for responses

Inferential Stats
```{r}
library(lme4)
library(lmerTest)
unique(analysis_data$Valence)
analysis_data$Valence <- as.factor(analysis_data$Valence)
levels(analysis_data$Valence)
contrasts(analysis_data$Valence) <- c(-.5, .5) #negative valence coded as negative

#try mixed models, if that fails use regressions

### White models

white_response_model1 <- glmer(white~Valence*Participant_Race + (1|Participant)+(1|Items),
family="binomial", analysis_data)

#singular

summary(white_response_model1)

white_response_model2 <- glmer(white~Valence*Participant_Race + (1|Items),
family="binomial", analysis_data)

summary(white_response_model2)

#white model --> people selected white for negative valence items in general
#white model --> trending towards choosing white more for black and white people compared to
'other', and trending interaction (doesn't matter, not significant)

### Black models

black_response_model1 <- glmer(black~Valence*Participant_Race + (1|Items),
family="binomial", analysis_data)

#doesn't work, need to use regression

black_response_reg_model1 <- lm(black~Valence*Participant_Race, data=analysis_data)

summary(black_response_reg_model1)

#black model --> people choose black more for positive valence items
#interactions
```

````
str(analysis_data)
#remove NAs from responses
intplot_data<- analysis_data[!(is.na(analysis_data$black) | analysis_data$black==""), ]

interaction.plot(intplot_data$Valence, intplot_data$Participant_Race, intplot_data$black)

#black model --> other race participants have a bigger bias towards choosing white for positive valence items


### Latina model

latina_response_reg_model1 <- lm(latina~Valence*Participant_Race, data=analysis_data)

summary(latina_response_reg_model1)

intplot_data<- analysis_data[!(is.na(analysis_data$latina) | analysis_data$latina==""), ]

interaction.plot(intplot_data$Valence, intplot_data$Participant_Race, intplot_data$latina)

#latina model --> white people select latina for positive valence items, black and other participants select latina for negative valence items

```


### Graph Response Data
```{r}
library(gridExtra)

unique(clean_survey$Race)
kiel_graph <- clean_survey

#remove nas
kiel_graph <- kiel_graph[!(is.na(kiel_graph$Response) | kiel_graph$Response==""), ]


#Positive graphs
kiel_white <- ddply(kiel_graph[kiel_graph$Race=="White",], .())
kiel_white_pos <- ddply(kiel_white[kiel_white$Valence=="1",], .())
kiel_white_neg <- ddply(kiel_white[kiel_white$Valence=="0",], .())

kiel_white_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_white_pos)) -> kiel_white_pos_perc
````

```
kiel_white_pos_graph <- ggplot(kiel_white_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("White Positive")
  #ylim(0,.55)

kiel_white_pos_graph

kiel_black <- ddply(kiel_graph[kiel_graph$Race=="Black",], .())
kiel_black_pos <- ddply(kiel_black[kiel_black$Valence=="1",], .())
kiel_black_neg <- ddply(kiel_black[kiel_black$Valence=="0",], .())

kiel_black_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_black_pos)) -> kiel_black_pos_perc

kiel_black_pos_graph <- ggplot(kiel_black_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Black Positive")+
  ylim(0,.55)

kiel_black_pos_graph

kiel_other <- ddply(kiel_graph[kiel_graph$Race=="Other",], .())
kiel_other_pos <- ddply(kiel_other[kiel_other$Valence=="1",], .())
kiel_other_neg <- ddply(kiel_other[kiel_other$Valence=="0",], .())

kiel_other_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_other_pos)) -> kiel_other_pos_perc

kiel_other_pos_graph <- ggplot(kiel_other_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Other Positive")
  #ylim(0,.55)

kiel_other_pos_graph

#Negative Graphs
kiel_white_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_white_neg)) -> kiel_white_neg_perc

kiel_white_neg_graph <- ggplot(kiel_white_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("White Negative")+
  ylim(0,.55)
```

````
kiel_white_neg_graph

kiel_black_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_black_neg)) -> kiel_black_neg_perc

kiel_black_neg_graph <- ggplot(kiel_black_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Black Negative")+
  ylim(0,.55)

kiel_black_neg_graph

kiel_other_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_other_neg)) -> kiel_other_neg_perc

kiel_other_neg_graph <- ggplot(kiel_other_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Other Negative")+
  ylim(0,.55)

kiel_other_neg_graph

#combine graphs

kiel_graph <-
arrangeGrob(kiel_white_pos_graph,kiel_white_neg_graph,kiel_black_pos_graph,kiel_black_neg
_graph,kiel_other_pos_graph,kiel_other_neg_graph,nrow=3,top="Response Bias")

ggsave(file="ABC2_ResponseBias.tiff", dpi=300,scale=2, kiel_graph)

```



### Accuracy Analysis
```{r}
acc_analysis <- read.csv("ABC2_Accuracy.csv")

acc_analysis$Speaker <- acc_analysis$Item

acc_analysis$Speaker <- revalue(acc_analysis$Speaker, c("R1"="Black", "R2"="Black",
"R3"="Black", "R4"="Black", "R5"="Latina", "R6"="Latina", "R7"="Latina", "R8"="White",
"R9"="White", "R10"="White"))
unique(acc_analysis$Speaker)
````

part_race <- analysis_data %>% select(Participant,Participant_Race)

acc_merged <- merge(part_race,acc_analysis)
#First do accuracy
unique(acc_merged$Participant_Race)
#join hispanic and other
#dummy code participant race
acc_merged$Black_part <- ifelse(acc_merged$Participant_Race=="Black",1,0)
acc_merged$Other_part <- ifelse(acc_merged$Participant_Race=="Other",1,0)
#speaker
unique(acc_merged$Speaker)

mean(acc_merged$Accuracy)

large_acc_model <- glmer(Accuracy~Participant_Race*Speaker+(1|Participant)+(1|Items),
family="binomial", acc_merged)

summary(large_acc_model)

#black participants less accurate overall

interaction.plot(acc_merged$Speaker, acc_merged$Participant_Race, acc_merged$Accuracy)

#black worst for black voice, equally bad for latina voice
```

### Accuracy Plot

```{r}
#make the means dataframe
accuracy_means <- ddply(acc_merged, .(Participant_Race,Speaker), summarize,
mean=mean(Accuracy),sd=sd(Accuracy),nd = n_distinct(Participant))

library(ggthemes)

is.numeric(accuracy_means$mean)

acc_means_plot <- ggplot(accuracy_means,aes(y=mean,x=Participant_Race,bg=Speaker))+
  geom_bar(position="dodge",stat="identity")+
  geom_errorbar(aes(ymin=mean-sqrt((mean*(1-mean))/(nd-1)),ymax=mean+sqrt((mean*(1-
mean))/(nd-1)), width=.1), position=position_dodge(width=.9),size=.3)+
  ggtitle("Accuracy Means")+
  ylab("Mean Accuracy")+
  xlab("Participant Race")+
  theme_calc()

```
acc_means_plot

ggsave(
"ABC2_acc_means.tiff",
plot = acc_means_plot,
device = "tiff",
dpi = 300,
units = c("in"),
scale = 1,
limitsize=TRUE
)
```

Experiment 3

```
raw_survey <- read.csv("Survey.csv")
raw_accuracy <- read.csv("Accuracy (1).csv")

#need to remove bad surveys, put NA for all or all but one or two. the first one is mine i forgot to
delete lol

bad_survey <- c("a3e27ac8a127397ac53b8515110ad70f",
"dcb675912f3147095c04e4adae03f432", "54f211789074bc90713d3124e74940aa",
          "3e5d7a92f88f7cddd40728650d6eee6e", "605bdd5418afa2624407dc91622f8d19",
"735e5760570a4d43bf479e0ec84a8418",
          "998f585010d71c16b714018273294786", "902b59580e3abc19d0d6a899a1fa3d89",
"022ecf69636c75bacfef0b17bbd7709a",
          "91cb9dc36cbe33320dce69d9fd903bbf", "50a28558f84c2ef032c8137e8d117c7f",
"5327c0fc2db92234ff994c20fb2cfcc3",
          "0b8518701e204f3998cb2fb8a0ceb642", "db26a2f68704c04132b0b91c666a8007")

raw_accuracy$bad <- raw_accuracy$Participant %in% bad_survey

library(plyr)

clean_accuracy <- ddply(raw_accuracy[raw_accuracy$bad==FALSE,], .())

#let's look at the data format for each column
str(clean_accuracy)

mean(clean_accuracy$Accuracy) #74.46%

#means by participant to look for outlier participants

bypart_accmeans <- ddply(clean_accuracy, .(Participant), summarize, mean=mean(Accuracy))
hist(bypart_accmeans$mean)
```

```
#oh wow not bad

byitem_accmeans <- ddply(clean_accuracy, .(Item), summarize, mean=mean(Accuracy))
byitem_accmeans

#HOLY CRAP THEY DID SO GOOD, remove the people with less than 50% to follow protocol
though
remove_from_survey <- c("a3e27ac8a127397ac53b8515110ad70f",
"dcb675912f3147095c04e4adae03f432", "54f211789074bc90713d3124e74940aa",
                "3e5d7a92f88f7cddd40728650d6eee6e",
"605bdd5418afa2624407dc91622f8d19", "735e5760570a4d43bf479e0ec84a8418",
                "998f585010d71c16b714018273294786",
"902b59580e3abc19d0d6a899a1fa3d89", "022ecf69636c75bacfef0b17bbd7709a",
                "91cb9dc36cbe33320dce69d9fd903bbf", "50a28558f84c2ef032c8137e8d117c7f",
"5327c0fc2db92234ff994c20fb2cfcc3",
                "0b8518701e204f3998cb2fb8a0ceb642",
"db26a2f68704c04132b0b91c666a8007", "0a47f8f7210060eb581e17e2b7b4a4b3",
                "1aa16558a71dbf1ec7ef6de20e77bb8d",
"c5e8d1ebae4141579a1146de6c55224a")

library(tidyverse)

#Survey Data**

raw_survey$bad <- raw_survey$Participant %in% remove_from_survey

clean_survey <- ddply(raw_survey[raw_survey$bad==FALSE,], .())

view(clean_survey)


unique(clean_survey$Participant) #61 participants
#survey type
str(clean_survey)
pos_item <- c("Survey1","Survey4","Survey5","Survey8")

clean_survey$Valence_bool <- clean_survey$Item %in% pos_item
clean_survey$Valence <- ifelse(clean_survey$Valence_bool=="TRUE",1,0)

#participant race
str(clean_survey)
unique(clean_survey$Race)

#organize other
clean_survey$Race <- revalue(clean_survey$Race, c("N/A"="Other", "Hispanic"= "Other",
"Native American"="Other", "Asian"="Other"))
```

````
unique(clean_survey$Race)

#now make Other the baseline for participant race

clean_survey$Participant_Race <- as.factor(clean_survey$Race)

summary(clean_survey$Participant_Race)

clean_survey$Participant_Race <- factor(clean_survey$Participant_Race,
c("Other","White","Black"))

summary(as.factor(clean_survey$Response))

clean_survey$white <- ifelse(clean_survey$Response=="White", 1,0)
clean_survey$black <- ifelse(clean_survey$Response=="Black", 1,0)
clean_survey$latina <- ifelse(clean_survey$Response=="Latina", 1, 0)

#now delete unnecessary columns

analysis_data <- clean_survey %>% select(Participant, Item, Age, Gender, Participant_Race,
Valence, white, black, latina)



#We need columns for: Participant, Age, Part_Gender, Part_Race, Item, Response, Valence, and
3 dummy variables for responses

#Inferential Stats
```{r}
library(lme4)
library(lmerTest)
citation("lme4")
unique(analysis_data$Valence)
analysis_data$Valence <- as.factor(analysis_data$Valence)
levels(analysis_data$Valence)
contrasts(analysis_data$Valence) <- c(-.5, .5) #negative valence coded as negative

#try mixed models, if that fails use regressions

### White models
levels(analysis_data$Participant_Race)
white_response_model1 <- glmer(white~Valence*Participant_Race + (1|Participant)+(1|Item),
family="binomial", analysis_data)

#singular
````
````

```
summary(white_response_model1)

white_response_model2 <- glmer(white~Valence*Participant_Race + (1|Item),
family="binomial", analysis_data)

summary(white_response_model2)


### Black models

black_response_model1 <- glmer(black~Valence*Participant_Race + (1|Item),
family="binomial", analysis_data)

#doesn't work, need to use regression

black_response_reg_model1 <- lm(black~Valence*Participant_Race, data=analysis_data)

summary(black_response_reg_model1)


#interactions
str(analysis_data)
#remove NAs from responses
intplot_data<- analysis_data[!(is.na(analysis_data$black) | analysis_data$black==""), ]

interaction.plot(intplot_data$Valence, intplot_data$Participant_Race, intplot_data$black)

#black model --> other race participants have a bigger bias towards choosing white for positive
valence items


### Latina model

latina_response_reg_model1 <- lm(latina~Valence*Participant_Race, data=analysis_data)

summary(latina_response_reg_model1)

intplot_data<- analysis_data[!(is.na(analysis_data$latina) | analysis_data$latina==""), ]

interaction.plot(intplot_data$Valence, intplot_data$Participant_Race, intplot_data$latina)

#latina model --> white people select latina for positive valence items, black and other
participants select latina for negative valence items

```

```
### Graph Response Data
```{r}
library(gridExtra)

unique(clean_survey$Race)
kiel_graph <- clean_survey

#remove nas
kiel_graph <- kiel_graph[!(is.na(kiel_graph$Response) | kiel_graph$Response==""), ]


#Positive graphs
kiel_white <- ddply(kiel_graph[kiel_graph$Race=="White",], .())
kiel_white_pos <- ddply(kiel_white[kiel_white$Valence=="1",], .())
kiel_white_neg <- ddply(kiel_white[kiel_white$Valence=="0",], .())

kiel_white_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_white_pos)) -> kiel_white_pos_perc

kiel_white_pos_graph <- ggplot(kiel_white_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("White Positive")
#ylim(0,.55)

kiel_white_pos_graph

kiel_black <- ddply(kiel_graph[kiel_graph$Race=="Black",], .())
kiel_black_pos <- ddply(kiel_black[kiel_black$Valence=="1",], .())
kiel_black_neg <- ddply(kiel_black[kiel_black$Valence=="0",], .())

kiel_black_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_black_pos)) -> kiel_black_pos_perc

kiel_black_pos_graph <- ggplot(kiel_black_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Black Positive")+
  ylim(0,.55)

kiel_black_pos_graph

kiel_other <- ddply(kiel_graph[kiel_graph$Race=="Other",], .())
kiel_other_pos <- ddply(kiel_other[kiel_other$Valence=="1",], .())
kiel_other_neg <- ddply(kiel_other[kiel_other$Valence=="0",], .())
```
```

```
kiel_other_pos %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_other_pos)) -> kiel_other_pos_perc

kiel_other_pos_graph <- ggplot(kiel_other_pos_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Other Positive")
#ylim(0,.55)

kiel_other_pos_graph

#Negative Graphs
kiel_white_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_white_neg)) -> kiel_white_neg_perc

kiel_white_neg_graph <- ggplot(kiel_white_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("White Negative")+
  ylim(0,.55)

kiel_white_neg_graph

kiel_black_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_black_neg)) -> kiel_black_neg_perc

kiel_black_neg_graph <- ggplot(kiel_black_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Black Negative")+
  ylim(0,.55)

kiel_black_neg_graph

kiel_other_neg %>%
  count(Response) %>%
  mutate(perc = n / nrow(kiel_other_neg)) -> kiel_other_neg_perc

kiel_other_neg_graph <- ggplot(kiel_other_neg_perc, aes(x=Response,y=perc))+
  geom_bar(stat="identity", position="dodge")+
  ggtitle("Other Negative")+
  ylim(0,.55)

kiel_other_neg_graph

#combine graphs
```

````
kiel_graph <-
arrangeGrob(kiel_white_pos_graph,kiel_white_neg_graph,kiel_black_pos_graph,kiel_black_neg
_graph,kiel_other_pos_graph,kiel_other_neg_graph,nrow=3,top="Response Bias")

ggsave(file="ABC3_ResponseBias.tiff", dpi=300,scale=2, kiel_graph)

```


### Accuracy Analysis
```{r}
acc_analysis <- read.csv("Accuracy (1).csv")

acc_analysis$Speaker <- acc_analysis$Item

acc_analysis$Speaker <- revalue(acc_analysis$Speaker, c("R1"="Latina", "R2"="Latina",
"R3"="Latina", "R4"="Latina", "R5"="White", "R6"="White", "R7"="White", "R8"="Black",
"R9"="Black", "R10"="Black"))
unique(acc_analysis$Speaker)

part_race <- analysis_data %>% select(Participant,Participant_Race)

acc_merged <- merge(part_race,acc_analysis)
#First do accuracy
unique(acc_merged$Participant_Race)
#join hispanic and other
#dummy code participant race
acc_merged$Black_part <- ifelse(acc_merged$Participant_Race=="Black",1,0)
acc_merged$Other_part <- ifelse(acc_merged$Participant_Race=="Other",1,0)
#speaker
unique(acc_merged$Speaker)

mean(acc_merged$Accuracy) #78

large_acc_model <- glmer(Accuracy~Participant_Race*Speaker+(1|Participant)+(1|Items),
family="binomial", acc_merged)

summary(large_acc_model)

interaction.plot(acc_merged$Speaker, acc_merged$Participant_Race, acc_merged$Accuracy)


```


### Accuracy Plot
````
````

````
```{r}
#make the means dataframe
accuracy_means <- ddply(acc_merged, .(Participant_Race,Speaker), summarize,
mean=mean(Accuracy),sd=sd(Accuracy),nd = n_distinct(Participant))

library(ggthemes)

is.numeric(accuracy_means$mean)

acc_means_plot <- ggplot(accuracy_means,aes(y=mean,x=Participant_Race,bg=Speaker))+
  geom_bar(position="dodge",stat="identity")+
  geom_errorbar(aes(ymin=mean-sqrt((mean*(1-mean))/(nd-1)),ymax=mean+sqrt((mean*(1-
mean))/(nd-1)), width=.1), position=position_dodge(width=.9),size=.3)+
  ggtitle("Accuracy Means")+
  ylab("Mean Accuracy")+
  xlab("Participant Race")+
  theme_calc()

acc_means_plot

ggsave(
  "ABC3_acc.tiff",
  plot = acc_means_plot,
  device = "tiff",
  dpi = 300,
  units = c("in"),
  scale = 1,
  limitsize=TRUE
)
````
````

APPENDIX C: IRB LETTER

ILLINOIS

OFFICE OF THE VICE CHANCELLOR
FOR RESEARCH & INNOVATION

Office for the Protection of Research Subjects
805 W. Pennsylvania Ave., MC-095
Urbana, IL 61801-4822

Notice of Approval: New Submission

February 6, 2020

Principal Investigator	Kiel Christianson
CC	Victoria Susberry
Protocol Title	*Accent Bias in Comprehension (ABC, versions 1, 2, 3 and Recording)*
Protocol Number	20500
Funding Source	Unfunded
Review Type	Expedited 6, 7
Status	Active
Risk Determination	No more than minimal risk
Approval Date	February 6, 2020
Closure Date	February 5, 2025

This letter authorizes the use of human subjects in the above protocol. The University of Illinois at Urbana-Champaign Institutional Review Board (IRB) has reviewed and approved the research study as described.

The Principal Investigator of this study is responsible for:

- Conducting research in a manner consistent with the requirements of the University and federal regulations found at 45 CFR 46.
- Using the approved consent documents, with the footer, from this approved package.
- Requesting approval from the IRB prior to implementing modifications.
- Notifying OPRS of any problems involving human subjects, including unanticipated events, participant complaints, or protocol deviations.
- Notifying OPRS of the completion of the study.