THE UNIVERSITY OF CHICAGO

WHAT'S IN A VOICE? EFFECTS OF DIALECT PERCEPTION ON ACTIVATION OF CRIME STEREOTYPES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

REBECCA KALA ROSEN

CHICAGO, ILLINOIS

AUGUST 2017

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Many thanks to my advisor, Daniel Casasanto, for his willingness to explore this research, for opening his lab to me, and for consistently advocating on my behalf. He and Laura Staum Casasanto have been patient and supportive collaborators who have pushed me to be a better scientist. I'd like to thank my committee members, Jennifer Kubota and Jasmin Cloutier, for their expertise, suggestions, and use of their lab resources during this process.

This research would not have been possible without the support of the Casasanto Experience and Cognition Lab members. They have been especially critical in considering the appropriate statistical analyses for the studies outlined here. Thank you to the Kenny Lab for its lively discussions, methodological training, welcoming attitude, and for sparking an interest in so many areas for me during my first years as a graduate student. I'm thankful to many friends in the department, who acted as a sounding board for my ideas and provided distractions when needed.

I am indebted to my family, who have been supportive and cheered me on, and who believed I could accomplish this from the start. I am thankful to them for limiting their requests to know when I would finish. I'd especially like to thank Jonny Ramos, who accompanied me on this journey.

Finally, it is my hope that with this research, students from all backgrounds will know that there is a place for them and their interests in higher education.

Abstract

Prior research has shown that after viewing Black males faces, participants are faster at categorizing weapons and slower at categorizing tools than after viewing White male faces. The present research seeks to extend this effect to determine whether hearing African American Vernacular English (AAVE) elicits these same threat-related stereotypes. In Study 1, we replicate a previous finding using a national online sample. In Studies 2a and 2b we establish a paradigm by which to examine the effects of dialect on the automatic activation of crime stereotypes. We found that hearing AAVE activates stereotypes, leading to faster categorization of weapons and slower categorization of tools than hearing Standard American English. We discuss the possibility that these effects may be especially notable in visually impoverished environments when auditory cues are the strongest indicator of race and the implications for AAVE speakers.

Introduction

Imagine that you are awoken by male voices shouting outside your urban building late at night, but you can't quite make out what they are saying. Should you assume that they are just inconsiderate partiers coming home late and go back to bed? Or should you be concerned that this is an altercation that could escalate and call the police? Without seeing their behavior, the one thing you have to go on is their voices. Are those voices speaking Standard American English (SAE) or African American Vernacular English (AAVE)? Perhaps their dialect will influence how you perceive the event and the actions you take.

To address this issue, we began by replicating previous findings that viewing Black males faces leads to faster categorization of weapons and slower categorization of tools than after viewing White male faces using the Weapons Identification Task (WIT; Payne, 2001). We replicated this finding with a national online sample using Amazon Mechanical Turk and slightly modified the WIT to rule out alternative explanations. Once we established that we had found a task that could be administered and replicated with our online sample, we modified the task to present participants with voice samples in AAVE and SAE dialects in order to address our question of interest.

Implicit Racial Bias

It has been well established that our judgments are often biased by race. Payne (2001) found that when asked to categorize guns and tools, participants were faster at categorizing guns after being primed with Black male faces than after being primed with White male faces, and faster at categorizing tools after being primed with White male faces than after being primed with Black male faces (Experiment 1). Further, when prompted to respond quickly, participants were more likely to miscategorize tools as weapons after being primed with Black male faces than after being primed with White male faces (Experiment 2). Eberhardt and colleagues (2004) have found that after being subliminally primed with Black male faces (as opposed to White male faces), White people are quicker to identify dangerous objects, such as guns, in degraded images. Researchers suggest that these effects arise because Black males are commonly stereotyped in the United States as being dangerous or aggressive. When Black male faces are primed, participants are anticipating dangerous objects and are therefore faster to identify weapons or more likely to falsely believe they've seen a weapon.

Activation of the crime stereotype happens regardless of whether the viewer is themselves high- or low-prejudiced (Devine, 1989). This implicit cognition is characterized by "past experience [that] influences judgment in a fashion not introspectively known by the actor" (Greenwald & Banaji, 1995, p. 4). For example, subliminally priming 'black' facilitates a lexical decision on negatively valenced words that are stereotypic of African Americans, whereas priming 'white' facilitates lexical decision on positively valenced words that are stereotypic of White Americans (Wittenbrink, Judd, & Park, 1997). Stereotype activation can serve to preserve resources (Macrae, Milne, & Bodenhausen, 1994) and fatigue can increase implicit racial bias (Ma et al., 2013).

It is clear, then, that when White people encounter Black male faces, negative stereotypes of Blacks are often activated. This phenomenon has the potential to lead to fatal outcomes for Black men. Indeed, there are documented cases of Black men who have been shot by police or others who have mistakenly believed that they were holding a weapon or otherwise posing a threat. Implicit measures have been shown to predict behavioral, judgment, and physiological measures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), and have correlation effect sizes that are large enough to explain discriminatory impacts (Greenwald, Banaji, & Nosek, 2015).

Correll and colleagues (2002) investigated how the race of the target affects shooters' decisions and found that novice shooters are in fact slower (Experiment 1) and less accurate (Experiment 2) when deciding not to shoot unarmed Black male targets than unarmed White male targets during a simulation, though bias is attenuated in trained police officers (Correll et al., 2007). (See Pearson, Dovidio, & Gaertner, 2009 for further review of implicit bias and aversive racism.)

Dialect as a Racial Cue

In much of the research described above on implicit racial bias, race is visually primed with pictures of Black and White individuals. Other research has primed the category of a race or nationality using words associated with that culture (e.g., Mange, Shavrit, Margas, & Sénémeaud, 2016). Our interest is whether auditory dialect cues can be used to prime implicit biases in the same way.

Research has shown that listeners are able to accurately match three dialects—Standard American English, African American Vernacular English, and Chicano English— to their associated races after hearing an example of that dialect, and that a speaker's dialect can be identified from a single word, such as "hello" (Purnell, Idsardi, & Baugh, 1999). Furthermore, these researchers demonstrated that dialect alone was sufficient to lead to housing discrimination. Therefore, we know that not only can even a short utterance lead to racial identification, it can also lead to prejudiced behaviors. In another study (Campbell-Kibler, 2012), characteristics of Southern English were implicitly associated with stereotypes of Southern states, such as blue collar workers. Similarly, 9- and 10-year old children have been shown to endorse stereotypes of Northern and Southern Americans when characterizing a Northern- or Southern-accented speaker (Kinzler & DeJesus, 2013).

It is clear, then, that listeners can rely on dialect to ascertain information about a speaker's race, and have used dialect to generate stereotypes about a speaker. Can dialect—specifically African American Vernacular English—also activate implicit stereotypes?

African American Vernacular English

African American Vernacular English (AAVE)—sometimes referred to as Black English or Ebonics—is a dialect of English spoken by some, though not all, Black Americans. It is characterized by distinctive vocabulary, pronunciations, and morphosyntax. While AAVE is spoken by many Americans and celebrated by many writers and researchers, it has been denigrated by others as unsophisticated and mischaracterized as simply incorrect English that lacks rules and grammar (Rickford & Rickford, 2000). One notable example is the controversy surrounding a 1996 resolution by the Oakland School Board to consider 'Ebonics' a distinct language that is spoken by some students. While some people applauded this decision others criticized it, and it became clear that at least some opponents considered AAVE to be a sign of limited education and held negative views of both the dialect and its speakers.

For the purposes of this study, AAVE stands in contrast to Standard American English (SAE): the English characteristic of academic instruction and textbooks in the United States. SAE is the dialect that is largely emphasized by dictionaries, grammar books, and standardized tests in the United States. While SAE and AAVE are considered separate dialects, they are spoken by overlapping populations; that is, many speakers of AAVE are bi-dialectal and speak SAE fluently. Furthermore, while some vocabulary, grammar, and pronunciation may differ between AAVE and SAE, they share features that are characteristic of many dialects of English.

Although there are myriad sociolinguistic studies on the usage of AAVE, far less attention has been paid to this dialect in psychology. Here we seek to determine whether hearing AAVE activates implicit racial stereotypes.

Current Research

As previously described, the finding that viewing Black males faces often activates crime stereotypes is well established. However, we know less about the judgments that are made about voices that are presumed to come from Black speakers. Does hearing AAVE—as opposed to SAE—elicit the same threat stereotype activation as being primed with Black faces? In Study 1, we replicate the known effect of threat stereotype activation when viewing Black male faces using a national online sample and we establish a novel protocol and method of analysis for this sample. In Studies 2a and 2b, we extend this effect and demonstrate that hearing AAVE automatically activates threat stereotypes in a similar manner.

Study 1: An online version of the Weapons Identification Task (WIT)

Payne (2001) demonstrated that when asked to categorize guns and tools, participants were faster at categorizing guns after being primed with Black male faces than after being primed with White male faces, and faster at categorizing tools after being primed with White male faces than after being primed with Black male faces using a Weapons Identification Task (WIT). In Study 1, we seek to replicate this finding within a national online sample in order to establish a paradigm and method of analysis that can later be extended to test the effect of dialect on activation of crime stereotypes.

Method

Participants

Participants were recruited for this study on Amazon Mechanical Turk (mturk) for monetary compensation. There were no restrictions placed on gender or race in order to encourage participants to report their demographics accurately; the only requirement was that participants must be located in the United States. Our study was accessed 1,045 times. We ran the study with an intended cut-off point of 1,000 participants on mturk, but the study was accessed an additional 45 times by people who did not indicate that they had completed it on mturk. This occurs if a participant signs up for the study on mturk but does not indicate that they have completed the study within the allotted time, or 'returns' the study after deciding not to complete it, but has already begun the procedure. We were left with 1,045 participants. 201 participants did not complete the study in its entirety and therefore could not be included in the analyses. Of the remaining participants, 15 participants were excluded from the analyses because their IP addresses or user identifications were the same as another participant's, and the veracity of the account could not be verified. An additional 16 participants were excluded from the analyses because they accessed the study from outside of the United States. In total, 813 participants' data were included in the statistical analyses.

Materials and Procedure

During this study, participants completed the Weapons Identification Task (WIT; Payne, 2001). During the WIT, participants are told that they will first view a picture of a face, and then either a weapon or a tool. Participants are told to ignore the face, which simply serves as a warning that the weapon or tool is about to appear. They are told to determine whether the second image is of a weapon or a tool. Half of the participants were told to press the 'E' key on the keyboard if the image is of a weapon, or the ‘I’ key on the keyboard if the image is of a tool; for the other half of participants, these instructions were reversed to counterbalance which key and hand would be used to categorize the target object. Such left–right counterbalancing is not discussed in the original publication of the WIT, but we included it in our procedure in order to rule out alternative explanations regarding the marked categories being associated with the left response hand or left side of the screen.

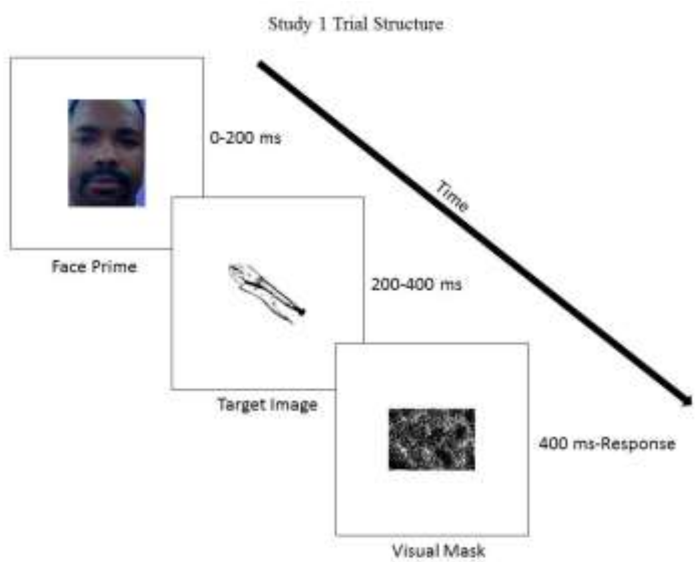
During this task, participants first viewed an image of either a Black male face, a White male face, or a neutral grey square for 200 ms in the center of a black screen, then viewed an image of either a weapon or a tool for 200 ms. An image mask (a black and white image of scrambled pixels) then appeared, replacing the weapon or tool, until participants pressed either the ‘E’ or ‘I’ key on the keyboard (see Figure 1).

The face primes and target images were all obtained from Keith Payne’s laboratory web page (Payne, n.d.). All of the images of black and white males were cropped to rectangles that only include the face of the male. The guns and tools are surrounded by a white, rectangular background.

The prime consisted of one of either 12 Black male faces or 12 White male faces, or a neutral grey box, that were presented at random in a within-subjects design. The target consisted of one of either 6 guns or 6 tools, which were also presented at random in a within-subjects design. Therefore, each trial consisted of a face or neutral prime and a target. Participants completed 12 practice trials, which were not scored, then 192 test trials, consistent with previous research using the WIT. Reaction times (RTs) were recorded, as well as whether the response was correct or incorrect.

Participants first signed up for the task on mturk. At this point, participants were only told that they would complete a study on categorization and perceptions. Once they accepted the study on mturk, participants were directed to click on a link to a Qualtrics (2015) survey. Once at Qualtrics, participants were presented with and accepted an informed consent form and provided their mturk ID numbers. They were then directed to a link on Inquisit (2015), where they could download the Inquisit software to complete the WIT. By using Inquisit, we were able to record accurate RTs using an online platform. After completing the WIT on Inquisit, participants returned to Qualtrics to complete demographic questions and a debriefing.

Figure 1. Trial structure for study 1.



Results

Each of the 813 participants included in the analyses completed 192 scored trials, totaling 156,096 trials. In order to analyze RTs, we excluded any trials on which participants responded faster than 100 ms or slower than 1,000 ms. This was also done a priori in Payne's (2001) original study according to standards laid out in Fazio (1990). There were 10,330 trials outside these boundaries. Furthermore, we only included the trials on which participants correctly categorized the weapons and tools; there were 8,317 additional incorrect trials dropped. This left us with 137,641 trials included in our analyses of reaction times.

All statistical analyses (unless otherwise noted) were performed using linear mixed effects models (Baayen, Davidson, & Bates, 2008) in R (R Development Core Team, 2016) and used the R packages lme4 (Bates, Maechler, Bolker, & Walker, 2015).

The main mixed-effects models of the results included the following fixed effects: Face Prime: White male face, Black male face, or a neutral grey square and Target Type: Tool or Weapon. Random effects were included where appropriate and consisted of random intercepts for subjects and for prime items (i.e., the specific prime file used). The dependent variable for each of these models was participants' log-transformed RTs, indicating the time it took in milliseconds for participants to categorize the tool or weapon image once it was presented.

We tested whether the interaction between Face Prime and Target Type influenced participants' RTs. The method of analyses described above compares a model with this the term of interest to one without and conducts a Chi-square goodness of fit test. There was a significant main effect of Face Prime ($\chi^2(2) = 46.56$, $p < .001$, $\varphi = 0.24$) and Target Type ($\chi^2(1) = 1752.3$, $p < .001$, $\varphi = 1.47$). We ascertained that Face Prime and Target Type interacted to predict RTs ($\chi^2(2) = 98.66$, $p < .001$, $\varphi = 0.35$). We conducted a follow-up analysis comparing the Face Prime by Target Type interaction in only those trials on which the Face Prime was Black or White, thereby excluding the control trials. This 2 x 2 interaction was also significant ($\chi^2(1) = 52.29$, $p < .001$, $\varphi = 0.25$), and simple effects testing revealed this was driven by participants being faster at categorizing weapons ($\chi^2(1) = 5.13$, $p = 0.02$, $\varphi = 0.08$) and slower at categorizing tools ($\chi^2(1) = 27.65$, $p < .001$, $\varphi = 0.18$) after viewing Black male faces (see Figures 2 and 3).

In Payne's 2001 study, only non-Black participants were included, and other research in this area has done the same or only included White participants. To determine whether our effects were characteristic of only White participants or could be generalized to participants of all races and ethnicities, we looked at the three-way interaction between Face Prime, Target Type, and Participant Race, comparing only White, non-Hispanic participants to all other participants. This interaction was neither significant when including the control Face Prime trials ($\chi^2(2) = 1.66$, $p = .44$, $\varphi = 0.05$), nor when examining only the Black and White Face Prime trials ($\chi^2(1) = 0.18$, $p = .67$, $\varphi = 0.01$). The two-way interaction of interest also remained significant when controlling for participant race for both the 3 x 2 analysis ($\chi^2(2) = 64.62$, $p < .001$, $\varphi = 0.28$) and the 2 x 2 analysis ($\chi^2(1) = 36.69$, $p < .001$, $\varphi = 0.21$).

To fully examine our data when considering participant race, we ran our analyses again examining only White, non-Hispanic participants' data (n = 603), then again examining only non-White participants' data (n = 210). Among White participants, there was a significant main effect of Face Prime ($\chi^2(2) = 35.44$, $p < .001$, $\varphi = 0.21$) and Target Type ($\chi^2(1) = 1259.8$, $p < .001$, $\varphi = 1.24$). We found a significant interaction in the 3 x 2 analysis ($\chi^2(2) = 67.85$, $p < .001$, $\varphi = 0.34$), and in the 2 x 2 analysis involving only Black and White Face Primes ($\chi^2(1) = 38.36$, $p < .001$, $\varphi = 0.25$). Simple effects testing revealed that participants were slower at categorizing tools after viewing Black faces ($\chi^2(1) = 21.18$, $p < .001$, $\varphi = 0.19$), but only marginally faster at categorizing weapons after viewing Black faces ($\chi^2(1) = 3.24$, $p = 0.0718$, $\varphi = 0.07$). Among non-White participants, there was a significant main effect of Face Prime ($\chi^2(2) = 31.02$, $p < .001$, $\varphi = 0.20$) and Target Type ($\chi^2(1) = 491.79$, $p < .001$, $\varphi = 0.78$). We also found a significant interaction in the 3x2 analysis ($\chi^2(2) = 31.33$, $p < .001$, $\varphi = 0.39$), and again in the 2 x 2 analysis involving only Black and White Face Primes ($\chi^2(1) = 13.996$, $p < .001$, $\varphi = 0.26$). However, the simple effects were significant for tools ($\chi^2(1) = 16.73$, $p < .001$, $\varphi = 0.28$) but not weapons ($\chi^2(1) = 1.93$, $p = 0.1649$, $\varphi = 0.096$).

Figure 2. Reaction time (ms) to categorize tools and weapons after a face or neutral prime.

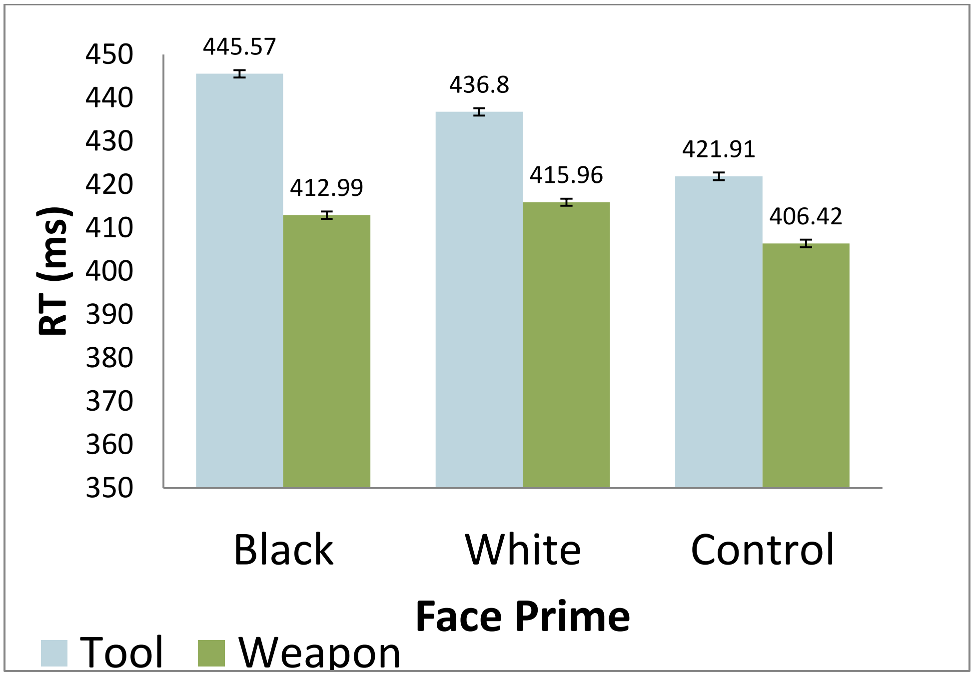
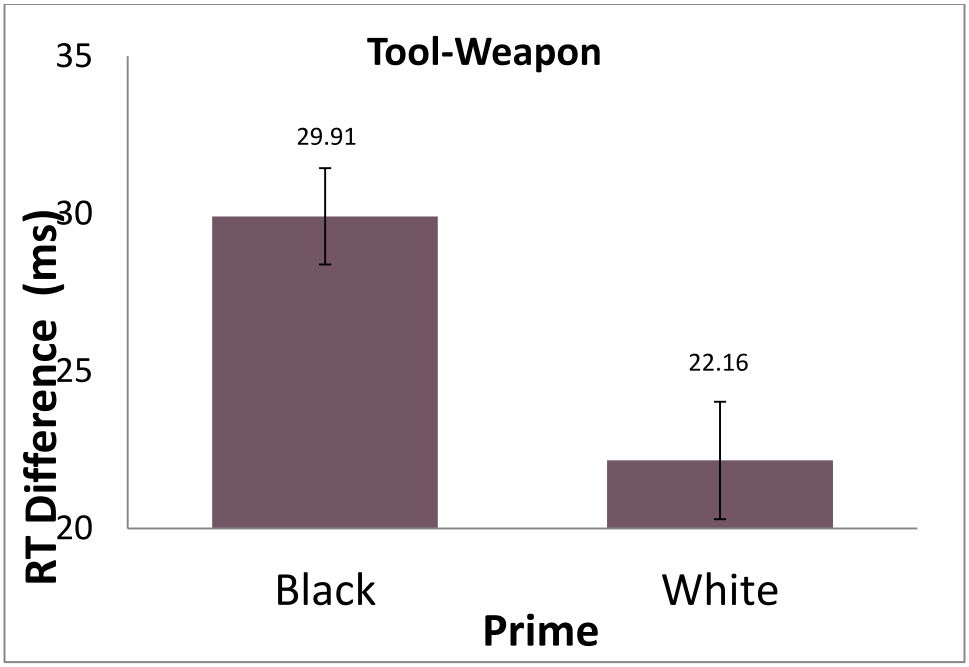


Figure 3. Difference score in reaction time (ms) to categorize tools and weapons as a function of face or neutral prime.



Discussion

In Study 1, we replicated previous findings that racial primes influence RTs when categorizing tools and weapons: participants were faster at categorizing weapons after viewing Black male faces than after viewing White male faces and were faster at categorizing tools after viewing White male faces than after viewing Black male faces. All interactions and simple effects of interest were significant in our whole sample, and in the White participants analyzed alone. The interactions of interest were also significant in the non-White participants alone. These results are consistent with previous research and inform us that the WIT effect can be replicated with our current paradigm and method of analysis and using a national online sample; we can now adapt this methodology to address our initial question of whether dialect primes automatically activate crime stereotypes. Further, by counterbalancing the side of the screen on which participants categorized tools and weapons, we can also rule out alternative explanations. We can now extend these findings to determine whether dialect primes elicit the same effects.

Study 2a: Does hearing a Black dialect automatically activate crime stereotypes?

In Study 2a, we seek to extend the effect of face primes on the automatic activation of stereotypes to determine whether voice primes also elicit threat-relevant racial stereotypes. For this study, we modified the WIT in order to present participants with auditory clips of Black males speaking AAVE and White males speaking SAE as our primes.

Method

Participants

Participants were recruited for this study on Amazon Mechanical Turk (mturk) for monetary compensation. There were no restrictions placed on gender or race in order to encourage participants to report their demographics accurately; the only requirement was that participants must be located in the United States. Our study was accessed 1,006 times. We ran the study with an intended cut-off point of 1,000 participants on mturk, but the study was accessed an additional 6 times by people who did not indicate that they had completed it on mturk, which left us with 1,006 participants. 140 participants did not complete the study in its entirety and therefore could not be included in the analyses. Of the remaining participants, 27 participants were excluded from the analyses because their IP addresses or user identifications were the same as another participant's, and the veracity of the account could not be verified. An additional 22 participants were excluded from the analyses because they accessed the study from outside of the United States. Because of the dual audio and visual nature of this study, participants were required to complete a manipulation check in which they were asked to indicate whether they recall any of the names stated by the voices; participants who did not list at least one correct name were excluded from the analyses. To this end, 235 participants were excluded from the analyses. While this points to a high manipulation check failure rate, it can be explained by the fact that many participants indicated they had technical audio difficulties and could not hear the sounds in the study; this may have had something to do with the participants' audio setups or operating systems, but is highly unlikely to correlate with any predictive measures that would influence participant performance on this task. In total, 582 participants' data were included in the statistical analyses.

Materials and Procedure

During the modified WIT, participants were told that they will first hear a sound clip of a voice, and then see a picture of either a weapon or a tool. Participants are told to ignore the voice, which simply serves as a warning that the weapon or tool is about to appear. They were told to determine whether the second image is of a weapon or a tool, and half of the participants were told to press the ‘E’ key on the keyboard if the image is of a weapon, or the ‘I’ key on the keyboard if the image is of a tool; for the other half of participants, these instructions were reversed to counterbalance which key and hand would be used to categorize the target object. During this task, participants first heard either a Black male voice stating a name in AAVE, a White male voice stating a name in SAE, or silence, then viewed either a weapon or tool for 200 ms. An image mask (a black and white image of scrambled pixels) then appeared until participants pressed either the ‘E’ or ‘I’ key on the keyboard (see Figure 4).

The target images were all obtained from Keith Payne’s laboratory web page (Payne, n.d.). The guns and tools are surrounded by a white, rectangular background. The prime consisted of one of either 12 Black male voices speaking AAVE or 12 White male voices speaking SAE, or an interval of no sound as a control condition that lasted the same duration as the longest of the voice clips, that were presented at random.

The voice samples that participants heard were drawn from a sample of 48 unique men—24 Black and 24 White—stating one of 24 names. To account for any effect of specific names, we counterbalanced which names were heard in AAVE and which were heard in SAE for each half of participants. The names used were generated to sound race neutral. Twenty of the given names used were from a list of crossover names—names equally likely to be from White and Black males—generated using California birth certificate information from 1961 to the year available at the time of the list’s creation (Levitt & Dubner, 2005). The other four given names were selected from a list of the most common names for babies with Black or White mothers in New York City in 2012. We selected four names from that list that were ranked within 1 point of commonness for Blacks and Whites; by using rank rather than raw numbers, we were able to account for the baseline number of Blacks and Whites in New York City. The surnames used were chosen by selecting surnames that were within 12 points in their rankings of commonness for self-identified White and Black people on the United States Census in 2000 (Butler, 2013). For example, 'Stewart' is the 46^{th} most common surname among Whites and 45^{th} most common surname among Blacks in the United States, and was selected on the basis of therefore being common among both Blacks and Whites. Ranking was used rather than the raw number of Whites and Blacks with the surname because Whites in the United States far outnumber Blacks; therefore, using raw numbers would have led to the selection of extremely uncommon surnames or surnames that are somewhat common for Blacks but uncommon for Whites. That is, ranking data allows us to control for the baseline numbers of Whites and Blacks in the United States.

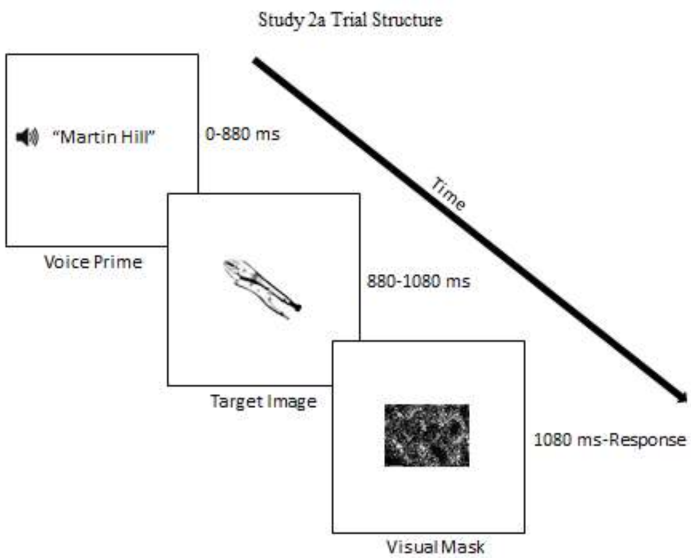
The 24 given names and surnames were then paired to produce 24 full names. While the names were paired pseudo-randomly, we switched any pairs that produced the full name of a reasonably famous person that participants may have recognized in order to eliminate distractions or confusion. We were left with 24 full names: George Stewart, Eric Evans, Dominic Parker, Cameron Taylor, Vincent Moore, Brandon Wilson, Shawn Davis, Frank Jones, Marc Williams, Gregory Johnson, Isaac Wright, Craig Reed, Troy Collins, Christian Young, Martin Hill, Corey Thomas, Micah Lewis, Nathaniel Walker, Aaron Allen, Theodore Smith, Julian Scott, Ryan King, Anthony Lee, and Andrew Thompson. We then recorded 24 White males who speak SAE and 24 Black males who speak AAVE stating one each of the names; the men were told to state the names in their normal speaking voices. By using names spoken by unique voices, we are providing individuating information about a person, analogous to the individuating information provided by a face in Study 1. However, this means that our AAVE and SAE samples vary on pronunciation and voice quality, but not on other aspects of dialect such as grammar and vocabulary.

Because the timing of the recordings varied, we added a pause to the beginning of each recording so that all of the recordings would last 880 ms, the duration of the longest recording. In this way, regardless of the sound clip presented on any particular trial, the target image could be timed to appear immediately after the sound clip finished playing.

Just as in Study 1, the target consisted of one of either 6 guns or 6 tools, which were presented at random in a within subjects design. Therefore, each trial consisted of a voice or neutral prime and a target. Participants completed 12 practice trials, which were not scored, then 192 test trials. Reaction times were recorded, as well as whether the response was correct or incorrect.

Participants signed up for the task on mturk. At this point, participants were only told that they would complete a study on categorization and perceptions. Once they accepted the study on mturk, participants were directed to click on a link to a Qualtrics survey. Once at Qualtrics, participants were presented with and accepted an informed consent form and provided their mturk ID numbers. They were then directed to a link on Inquisit, where they could download the Inquisit software to complete the WIT. Following the WIT, participants returned to Qualtrics to complete demographic questions and a debriefing.

Figure 4. Trial structure for study 2a.



Results

Each of the 582 participants included in the analyses completed 192 scored trials, totaling 111,744 trials. In order to analyze reaction times, we excluded any trials on which participants responded faster than 100 ms or slower than 1,000 ms. This was also done a priori in Payne's (2001) original study according to standards laid out by Fazio (1990). There were 4,307 trials outside these boundaries. Further, we only included the trials on which participants correctly categorized the weapons and tools; there were an additional 4,482 incorrect trials dropped, leaving us with 102,955 trials included in our analyses of reaction times.

All statistical analyses (unless otherwise noted) were performed using linear mixed effects models (Baayen, Davidson, & Bates, 2008) in R (R Development Core Team, 2016) and used the R packages lme4 (Bates, Maechler, Bolker, & Walker, 2015).

The main mixed-effects models of the results included the following fixed effects: Voice Prime: Standard American English (SAE), African American Vernacular English (AAVE), or time-matched silence and Target Type: Tool or Weapon. Random effects were included where appropriate and consisted of random intercepts for subject and for specific prime file used. The dependent variable for each of these models was participants' log-transformed reaction time (RT), or the time it took in milliseconds for participants to categorize the tool or weapon image once it was presented.

We tested whether the interaction between Voice Prime and Target Type influenced participants' RTs. The method of analyses described above compares a model with the term of interest to one without and conducts a Chi-square goodness of fit test. There was a significant main effect of Voice Prime ($\chi^2(2) = 45.41$, $p < .001$, $\varphi = 0.28$) and Target Type ($\chi^2(1) = 3312.8$, $p < .001$, $\varphi = 2.39$). We ascertained that Voice Prime and Target Type interacted to predict RTs ($\chi^2(2) = 12.17$, $p = .0023$, $\varphi = 0.14$). We conducted a follow-up analysis comparing the Voice Prime by Target Type interaction in only those trials on which the Voice Prime was SAE or AAVE, thereby excluding the control trials. This 2 x 2 interaction was also significant ($\chi^2(1) = 4.9002$, $p = .0269$, $\varphi = 0.09$). This was driven by a non-significant tendency for participants to categorize weapons faster after hearing AAVE ($\chi^2(1) = 1.83$, $p = .1764$, $\varphi = 0.06$) and categorize tools faster after hearing SAE ($\chi^2(1) = 3.25$, $p = .0715$, $\varphi = 0.07$).

In Payne's 2001 study, only non-Black participants were included, and other research in this area has done the same or only included White participants. To determine whether our effects were characteristic of only White participants or could be generalized to participants of all races and ethnicities, we looked at the three-way interaction between Face Prime, Target Type, and Participant Race, comparing only White, non-Hispanic participants to all other participants. This interaction was neither significant when including the control Face Prime trials ($\chi^2(2) = 2.39$, $p = .30$, $\varphi = 0.06$), nor when examining only the Black and White Face Prime trials ($\chi^2(1) = 0.59$, $p = .44$, $\varphi = 0.03$). However, the two-way interaction of interest became marginal when controlling for participant race for both the 3 x 2 analysis ($\chi^2(2) = 5.39$, $p = .07$, $\varphi = 0.096$) and the 2 x 2 analysis ($\chi^2(1) = 2.44$, $p = .12$, $\varphi = 0.06$).

To fully examine our data when considering participant race, we ran our analyses again examining only White, non-Hispanic participants' data (n = 444), then again examining only non-White participants' data (n = 138). Among White participants, there was a significant main effect of Voice Prime ($\chi^2(2) = 38.31$, $p < .001$, $\varphi = 0.26$) and Target Type ($\chi^2(1) = 2477$, $p < .001$, $\varphi = 12.06$). We found a marginal interaction in the 3 x 2 analysis ($\chi^2(2) = 5.38$, $p = .0678$, $\varphi = 0.11$), but not in the 2 x 2 analysis involving only SAE and AAVE Voice Primes ($\chi^2(1) = 2.44$, $p = .1186$, $\varphi = 0.07$). Among non-White participants, there was a significant main effect of Face Prime ($\chi^2(2) = 20.42$, $p < .001$, $\varphi = 0.19$) and Target Type ($\chi^2(1) = 838.71$, $p < .001$, $\varphi = 1.20$). We found a significant interaction in the 3 x 2 analysis ($\chi^2(2) = 9.21$, $p = .01$, $\varphi = 0.26$), but only a marginal interaction in the 2 x 2 analysis ($\chi^2(1) = 3.09$, $p = .0787$, $\varphi = 0.15$).

Discussion

Although the predicted interactions were significant in our whole sample, they became marginal when we controlled for participant race and only trended in the predicted direction in our analysis of White participants, alone. We therefore planned to replicate the study with a larger sample of participants, to test the same hypothesis, using the same task, with more statistical power. We conducted a post hoc power analysis to compute our achieved power in this study using G*Power 3.1 statistical software (Faul et al., 2007). For this power analysis, we used values from the model that included all participants, but only the specific contrast between SAE and AAVE Voice Primes on both weapons and tools (2 x 2 analysis). Given our effect size, $\varphi = 0.09$, n = 582, df = 1, and $\alpha = .05$, we computed an achieved power level of 58%. Because of this low level of power, we decided to conduct an additional study that would have greater power.

Study 2b: Replicating our effects of dialect on weapon identification with increased power

In Study 2a, not all of our analyses reached traditional levels of statistical significance. After conducting a power analysis, we determined that our achieved power in Study 2a was only 58%. We determined that doubling our sample size from 1,000 to 2,000 participants would give us a power level of at least 85%. This takes into account that in Study 2a, only 57.85% of our sample could be included in the analyses. Therefore, we expected that a gross sample size of 2,000 would yield a net usable sample of about n = 1,157. Using the values from Study 2a- effect size, $\varphi = 0.09$, df = 1, and $\alpha = .05$- a usable sample of 1,157 (netted from 2,000 participants in our overall sample) will yield a power level of 86.46%. For this reason, we ran Study 2b in exactly the same way as Study 2a, but with a sample cut-off of 2,000 participants.

Method

Participants

Participants were recruited for this study on Amazon Mechanical Turk (mturk) for monetary compensation. There were no restrictions placed on gender or race in order to encourage participants to report their demographics accurately; the only requirement was that participants must be located in the United States. Our study was accessed 2,067 times. We ran the study with an intended cut-off point of 2,000 participants on mturk, but the study was accessed an additional 67 times by people who did not indicate that they had completed it on mturk, which left us with 2,067 participants. 323 participants did not complete the study in its entirety and therefore could not be included in the analyses. Of the remaining participants, 354 participants were excluded from the analyses because their IP addresses or user identifications were the same as another participant's, and the veracity of the account could not be verified. An additional 60 participants were excluded from the analyses because they accessed the study from outside of the United States. Because of the dual audio and visual nature of this study, participants were required to complete a manipulation check in which they were asked to indicate whether they recall any of the names stated by the voices; participants who did not list at least one correct name were excluded from the analyses. To this end, 365 participants were excluded from the analyses. While this points to a high manipulation check failure rate, it can be explained by the fact that many participants indicated they had technical audio difficulties and could not hear the sounds in the study; this may have had something to do with the participants' audio setup, but is highly unlikely to correlate with any predictive measures that would influence participant performance on this task. In total, 965 participants' data were included in the statistical analyses.

Materials and Procedure

Study 2b was run using the exact same materials and procedures described in Study 2a. The only difference was the number of participants sampled, as described above.

Results

Each of the 965 participants included in the analyses completed 192 scored trials, totaling 185,280 trials. In order to analyze reaction times, we excluded any trials on which participants responded faster than 100 ms or slower than 1,000 ms. This was also done a priori in Payne's (2001) original study according to standards laid out by Fazio (1990). There were 7,910 trials outside these boundaries. Further, we only included the trials on which participants correctly categorized the weapons and tools; there were an additional 8,547 incorrect trials dropped, leaving us with 168,823 trials included in our analyses of reaction times.

All statistical analyses (unless otherwise noted) were performed using linear mixed effects models (Baayen, Davidson, & Bates, 2008) in R (R Development Core Team, 2016) and used the R packages lme4 (Bates, Maechler, Bolker, & Walker, 2015).

The main mixed-effects models of the results included the following fixed effects: Voice Prime: Standard American English (SAE), African American Vernacular English (AAVE), or time-matched silence and Target Type: Tool or Weapon. Random effects were included where appropriate and consisted of random intercepts for subject and for specific prime file used. The dependent variable for each of these models was participants' log-transformed reaction time (RT), or the time it took in milliseconds for participants to categorize the tool or weapon image once it was presented.

We tested whether the interaction between Voice Prime and Target Type influenced participants' RTs. The method of analyses described above compares a model with the term of ineterst to one without and conducts a Chi-square goodness of fit test. There was a significant main effect of Voice Prime ($\chi^2(2) = 29.65$, $p < .001$, $\varphi = 0.18$) and Target Type ($\chi^2(1) = 4902.6$, $p < .001$, $\varphi = 2.25$). We ascertained that Voice Prime and Target Type interacted to predict RTs ($\chi^2(2) = 9.00$, $p = .0111$, $\varphi = 0.097$). We conducted a follow-up analysis comparing the Voice Prime by Target Type interaction in only those trials on which the Voice Prime was SAE or AAVE, thereby excluding the control trials. This 2 x 2 interaction was also significant ($\chi^2(1) = 6.79$, $p = .0092$, $\varphi = 0.08$). This interaction was driven by a non-significant tendency for participants to categorize weapons faster after hearing AAVE ($\chi^2(1) = 0.78$, $p = .3775$, $\varphi = 0.03$) and to categorize tools significantly faster after hearing SAE ($\chi^2(1) = 5.77$, $p = .0163$, $\varphi = 0.08$; see Figures 5 and 6).

In Payne's 2001 study, only non-Black participants were included, and other research in this area has only included White participants. To determine whether our effects were characteristic of only White participants or could be generalized to participants of all races and ethnicities, we looked at the three-way interaction between Face Prime, Target Type, and Participant Race, comparing only White, non-Hispanic participants to all other participants. This interaction was neither significant when including the control Face Prime trials ($\chi^2(2) = 0.39$, $p = .82$, $\varphi = 0.02$), nor when examining only the Black and White Face Prime trials ($\chi^2(1) = 0.54$, $p = .46$, $\varphi = 0.02$). The two-way interaction of interest also remained significant when controlling for participant race for both the 3 x 2 analysis ($\chi^2(2) = 8.62$, $p = .01$, $\varphi = 0.09$) and the 2 x 2 analysis ($\chi^2(1) = 6.98$, $p = .008$, $\varphi = 0.09$).

To fully examine our data when considering participant race, we ran our analyses again examining only White, non-Hispanic participants' data (n = 742), then again examining only non-White participants' data (n = 223). Among White participants, there was a significant main effect of Voice Prime ($\chi^2(2) = 30.23$, $p < .001$, $\varphi = 0.18$) and Target Type ($\chi^2(1) = 3793$, $p < .001$, $\varphi = 1.98$). We found a significant interaction in the 3 x 2 analysis ($\chi^2(2) = 8.74$, $p = .0126$, $\varphi = 0.11$), and in the 2 x 2 analysis involving only SAE and AAVE Voice Primes ($\chi^2(1) = 7.09$, $p = .0078$, $\varphi = 0.098$). Whites were faster at categorizing tools after hearing SAE ($\chi^2(1) = 6.50$, $p = .0108$, $\varphi = 0.09$), but only marginally faster at categorizing weapons after hearing AAVE ($\chi^2(1) = 3.24$, $p = .0718$, $\varphi = 0.07$). Among non-White participants, there was a significant main effect of Face Prime ($\chi^2(2) = 13.32$, $p < .001$, $\varphi = 0.12$) and Target Type ($\chi^2(1) = 1110.3$, $p < .001$, $\varphi = 1.07$). we did not find a significant interaction in the 3 x 2 analysis ($\chi^2(2) = 0.84$, $p = .6584$, $\varphi = 0.06$), nor in the 2 x 2 analysis involving only AAVE and SAE Voice Primes ($\chi^2(1) = 0.3357$, $p =$

.5623, $\varphi = 0.04$). The simple effects were not significant for tools ($\chi^2(1) = 0.08$, $p = .7821$, $\varphi = 0.02$) nor for weapons ($\chi^2(1) = 0.32$, $p = .5708$, $\varphi = 0.04$).

Figure 5. Reaction time (ms) to categorize tools and weapons after a voice or neutral prime.

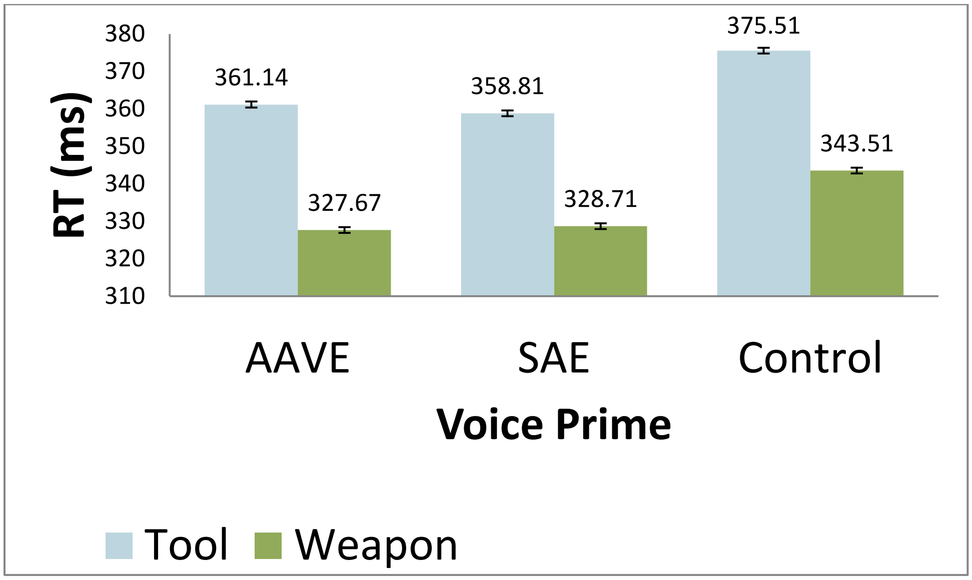
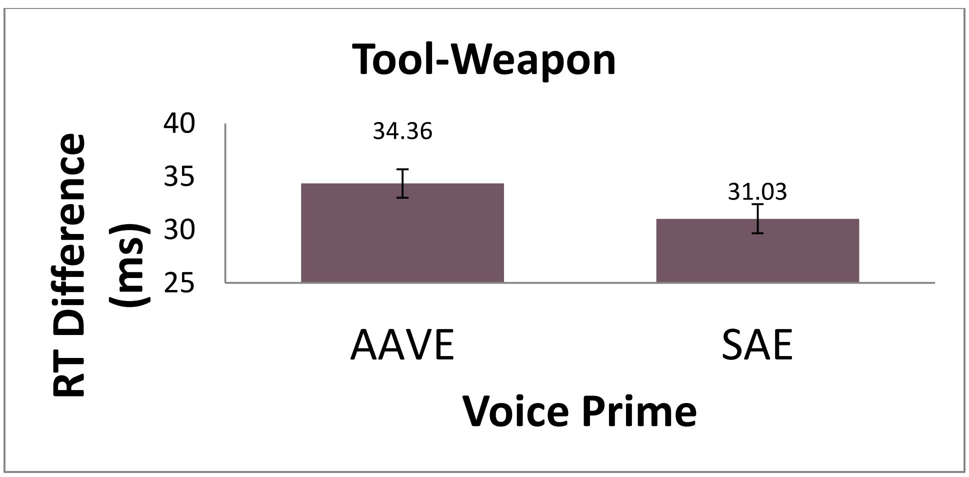


Figure 6. Difference score in reaction time (ms) to categorize tools and weapons as a function of voice or neutral prime.



General Discussion

In Study 1, we replicated the known effect that viewing Black males faces (vs. White males faces) leads to faster categorization of weapons and slower categorization of tools using a national online sample. In Studies 2a and 2b we established a paradigm by which we could extend the effect of face primes to dialect primes. We found that hearing Black males speaking AAVE (vs. White males speaking SAE) also leads to a tendency to categorize weapons faster and tools slower. While the effect size for the 2 x 2 analysis of the faces study ($\varphi = 0.26$) was larger than the effect sizes for the 2 x 2 analyses of voices ($\varphi = 0.09$ for Study 2a and $\varphi = 0.08$ for Study 2b), dialect did affect reaction time to categorize tools and weapons.

Previous research on AAVE has shown that Americans' knowledge of race-linked dialects can influence how they perceive speech and use information about dialect to resolve speech ambiguities. Staum Casasanto (2008) gave people faces and sentences, and found that viewing Black or White faces changed which word listeners thought they heard: when viewing Black faces, they were faster at discerning a sentence with an ambiguous word if it was consistent with a pronunciation characteristic of AAVE. So, while previous research has demonstrated that knowledge about AAVE can affect basic speech perception, the current research demonstrates that hearing AAVE can affect categorization of visual stimuli.

As described, each of our studies counterbalanced the side of the screen and hand with which tools and weapons were categorized. In Payne's (2001) study, counterbalancing is not described, meaning those data were left amenable to two alternative explanations beyond the assumed implicit racial bias mechanism: polarity alignment and implicit associations between space and valence. In the case of polarity alignment, items from a marked, or non-default, category are more closely associated with and processed more quickly on the left (Proctor & Cho, 2006); in this case, both weapons and Black males may be marked categories. Alternatively, it is possible that because an individual's non-dominant side is has a more negative valence (Casasanto, 2009)—and presumably most samples are comprised largely of right-handed participants—participants in previous studies were faster at categorizing Black males after viewing weapons because of an associated negativity with both categories. If this were the case, this still means that there is an implicit negativity about Black males, though it would not necessarily need to be specific to crime stereotypes. By counterbalancing the side on which weapons are categorized, we have ruled out the possible alternative explanations—an important distinction in isolating a mechanism for our observed results.

Returning to our original question, it seems then that African American Vernacular English activates crime stereotypes. While this was established using audio clips of males stating names, this does not take into account how more disparate examples of dialect may impact these results. Our samples do not contain dialect differences above and beyond accent alone; that is, while a mere difference in pronunciation of the same names was sufficient to elicit these results, voice samples that include differences in vocabulary and grammar that are typical of AAVE and SAE could magnify these effects.

Given that race can be inferred from even a single word spoken in dialect (Purnell et al., 1999), voice cues may be relied upon when other information is lacking. For example, a police officer making a shoot/don't shoot decision in a dark stairwell with obstructions may only be able to rely on auditory cues, including dialect. In some ways, dialect discrimination is one of the last allowable forms of racial discrimination. There are countless examples of well-respected individuals criticizing AAVE, whereas discrimination based on skin color is both illegal and taboo. Perhaps, then, provided with both dialect and visual cues of race, AAVE would be more heavily relied upon as an indicator of danger (regardless of whether AAVE speakers are actually dangerous). Therefore, although our effect sizes were larger for face primes than for voice primes, in a dark or otherwise visually impoverished environment voices may be more perceptible than faces, and dialect could serve as a stronger cue to race than skin color – and potentially a more potent activator of race-linked stereotypes. Furthermore, when simultaneously considering both visual and auditory race cues, AAVE may have a greater impact on the activation of stereotypes; this remains to be investigated.

Previous research has shown that the effects of race primes on reaction times in shoot/don't shoot decisions are attenuated in trained police officers (Correll et al., 2007), presumably because police officers are trained to attend to cues other than race in these scenarios. While this training has not completely eradicated the real-world incidence of unarmed individuals being erroneously shot—and other studies have shown that crime primes shift police officers' attention to Black faces (Eberhardt et al., 2004)—it is clear that training has attenuated these effects in the laboratory. While the studies described here are limited to categorization of weapons, future research should seek to understand the way in which dialect may influence shoot/don't shoot decisions, and how training to attend to cues other than dialect can reduce the potentially fatal outcomes for AAVE speakers who may incorrectly be associated with threat-relevant stereotypes. While many police training programs purport to train officers to reduce their attention to race in favor of true indicators of danger, less is known about how officers are trained to deal with dialect because dialect had not been shown previously to elicit these implicit biases.

It is important to consider how the targets of these biases, namely AAVE speakers, are impacted by these effects. Many bidialectal speakers engage in code-switching, or alternating between dialects, to produce speech they believe is situationally appropriate. In our design we had exclusively Black males speaking AAVE and exclusively White males speaking SAE. It is possible that hearing a Black-linked dialect is what elicited our observed effects, though it is also possible that simply hearing a Black speaker, regardless of dialect, would also elicit these effects.

If it is the latter, there may be racial differences in voice quality that are not specific to dialect that are sufficient to activate crime stereotypes. Future studies may investigate whether simply hearing a Black male can elicit these results regardless of whether he is speaking AAVE or SAE, and whether his level of fluency in SAE moderates these results.

There is limited evidence on how well multiple dialects can be produced by the same speakers; in one study, though, a linguist was able to produce 3 dialects convincingly enough that listeners to his voice samples guessed his race differently depending upon the dialect (Purnell et al., 1999). Given the potential repercussions of our results, a bidialectal AAVE speaker may choose to use SAE during certain interactions, such as during confrontations with police, in order to limit activation of negative racial stereotypes. Not every speaker of AAVE is bidialectal, though, and it is unclear how well a speaker can code-switch when experiencing a stressful interaction. Furthermore, if one's race can be perceived through cues such as skin color, this may affect how dialect is perceived if one's speech falls somewhere between AAVE and SAE. Niedzielski (1999) found that simply labeling a speaker as Canadian or from Detroit led to different perception of the speaker's vowels. Therefore, when other racial cues are present, switching one's dialect only slightly may not be sufficient to eradicate the crime stereotypes activated by speaking AAVE.

To what extent are our observed effects due to biases about AAVE itself, or did AAVE simply serve as a race prime? If that is the case, we would expect these effects to be strongest for dialect samples that can most reliably be identified as having been produced by a Black speaker, just as target prototypicality moderates racial bias in shoot/don't shoot decisions (Ma & Correll, 2011) and affective evaluations (Livingston & Brewer, 2002). Future research should seek to examine how prototypicality of the dialect sample moderates these effects.

Are the observed effects due to the dialect primes directly activating stereotypes about crime, or are these effects mediated by a fear response? In one study of WIT performance among military cadets, pupil dilation (a measure of autonomic arousal) predicted false positive rates. However, pupil dilation was significantly increased when viewing Middle Eastern males in traditional clothing, but not when viewing Black males, even though false positive error rates were high when viewing both of these groups (Fleming, Bandy, & Kimble, 2010). Therefore, while a fear or arousal response may contribute to some of these effects, it is not necessary for the effects to occur.

Finally, we note that while participant race did not interaction with our effects of interest, the effects observed in our analyses of non-White participants, alone, were not consistent across the three studies. The simplest explanation for the inconsistency is statistical power: our samples of non-White participants were much smaller than our samples of White participants, and therefore yielded less reliable estimates of the effect of race primes. One other possible explanation for this pattern, however, follows from a proposed mechanism by which stereotypes form. If both a group and a behavior are statistically infrequent, but are viewed to co-occur, observers may overestimate the frequency of co-occurrence (illusory correlation; Hamilton & Gifford, 1976). In this case, if an individual has few occasions to observe Black individuals, and relies solely on media portrayals of Blacks as violent or criminal, he may erroneously overestimate the relationship between Blacks and criminality. Because non-Whites may be more likely to have contact with other racial minority groups (e.g., U.S. public school have become increasingly segregated over the past few decades; Orfield, Ee, Frankenberg, & Siegel-Hawley, 2016), this may help explain why our effects were more consistent among White participants than among non-White participants. If this is the case, we would expect that increased lifetime contact with Blacks should attenuate these effects.

Despite these remaining questions, the current research fundamentally answers our original research question: does dialect automatically activate crime stereotypes? Yes; hearing dialect significantly impacted weapon and tool categorization.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48.

Butler, R. (2013). *Most common last names for Blacks in the U.S.* Retrieved from http://names.mongabay.com/data/black.html

Butler, R. (2013). *Most common last names for Whites in the U.S.* Retrieved from http://names.mongabay.com/data/white.html

Campbell-Kibler, K. (2012). The Implicit Association Test and sociolinguistic learning. *Lingua, 122*, 753–763.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General, 138,* 351–367.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83,* 1314–1329.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S. & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92,* 1006–1023.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Eberhardt, J., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87,* 876–893.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175–191.

Fazio (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology.* Newbury Park: Sage.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108,* 553–561.

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

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12,* 392–407.

Inquisit 4 [Computer software]. (2015). Retrieved from http://www.millisecond.com.

Kinzler, K. D., & DeJesus, J. M. (2013). Northern = smart and Southern = nice: The development of accent attitudes in the United States. *The Quarterly Journal of Experimental Psychology, 66,* 1146–1158.

Levitt, S. D., & Dubner, S. J. (April 11, 2005). *A Roshanda by Any Other Name: How do babies with super-black names fare?* Retrieved from http://pricetheory.uchicago.edu/levitt/Papers/ARoshandaByAnyOtherName.pdf

Livingston, R. W., & Brewer, M. B. (2002) What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82,* 5–18.

Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology, 47,* 391–396.

Ma, D. S., Correll, J., Wittenbrink, B., Bar-Anan, Y., Sriram, N., & Nosek, B. A. (2013). When fatigue turns deadly: The association between fatigue and racial bias in the decision to shoot. *Basic and Applied Social Psychology, 35,* 515–524.

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66,* 37–47.

Mange, J., Sharvit, K., Margas, N., & Sénémeaud, C. (2016). Do I shoot faster because I am thinking about an outgroup or a threatening outgroup? *Social Psychology, 47,* 29–37.

Monteith, M. J., Mark, A. Y., & Ashburn-Nardo, L. (2010). The self-regulation of prejudice: Toward understanding its lived character. *Group Processes & Intergroup Relations, 13,* 183–200.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18,* 62–85.

Orfield, G., Ee, J., Frankenberg, E., & Siegel-Hawley, G. (2016). *Brown at 62: School segregation by race, poverty, and state.* Retrieved from http://escholarship.org/uc/item/5ds6k0rd

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Attitudes and Social Cognition, 81,* 181–192.

Payne, B. K. (n.d.). *Weapon identification.* Retrieved from http://bkpayne.web.unc.edu/research-materials/

Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass, 3,* 314–338.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132,* 416–442.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18,* 10–30.

Qualtrics. (2015). Retrieved from http://www.qualtrics.com.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rickford, J. R., & Rickford, R. J. (2000). *Spoken soul: The story of Black English.* New York, NY: John Wiley & Sons.

Staum Casasanto, L. (2008). Does social information influence sentence processing? *Proceedings of the Annual Meeting of the Cognitive Science Society, 30,* 799–804.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72,* 262–274.

Appendix: Pilot Studies

Pilot studies were run that have not been reported in the main text because of flaws in the design. These studies were run prior to any of the studies described in the main text and are reported in full here.

Pilot Study A

In Pilot Study A, we hoped to provide a conceptual replication of Eberhardt and colleagues' (2004) study demonstrating faster weapon detection following priming of Black male faces. Rather than prime faces subliminally, however, we primed faces supraliminally. We did this in order to create a paradigm that presents faces in a manner that would match our presentation of voices. We predicted that viewing Black males faces would lead to faster weapon detection by White males in our study pool.

Method

Participants

Thirty-two White males in the University of Chicago community completed the study for either a $5 payment or partial completion of a course requirement. One participant was not included in the analyses because he had only lived in the United States for a year and was not able to provide all of the object names in English. This left 31 participants included in the analyses. All participants signed up for an experiment described as a study of audio and visual perception on the university's experiment website. Participants were randomly assigned to a 3 (race of prime: White prime ($n = 10$), Black prime ($n = 11$), or no-prime control ($n = 10$)) x 2 (object type: weapon or non-weapon) mixed-model design with race of prime serving as the between subject factor and object type serving as the within subject factor.

Materials

Participants in this study viewed several images of faces and then were asked to identify degraded images of objects. For the face images, 12 136 x 180 pixel images of either White or Black men's faces against a white background were presented consecutively in the center of a black screen for 500 ms each. The 12 images from each race were selected from the Eberhardt Face Database (Eberhardt, n.d.) so that the two race groups were matched as closely as possible on age, attractiveness, and stereotypicality. Participants in the control condition were not presented with any faces and were routed directly to the object detection task.

For the object detection task, we created 14 sets of object stimuli, using the same objects and method described by Eberhardt and colleagues (2004). Using Adobe Photoshop, we created a black and white line drawing of 4 weapons (2 guns and 2 knives) and 10 non-weapons (a book, a bugle horn, a camera, a key, a penny, a pocket watch, a staple remover, a stapler, a teacup and saucer, and a telephone). We then added "noise" to the image in equal increments in order to create 41 images ranging from extremely degraded to no noise. Participants were shown each image in the set, beginning with the most degraded image, for 500 ms each with no mask in between images. This gave the impression of a very degraded image of an object that slowly became more clear. Participants were instructed to press the spacebar as soon as they could accurately identify the object; upon pressing the spacebar the image would disappear and they were then prompted to type in the name of the object they saw. Once a response was entered, a fixation cross was presented, then the next set of object images was presented. The order in which the objects were presented was randomized without replacement.

Participants then responded to items from the Internal and External Motivation to Respond Without Prejudice Scales (IMS and EMS; Plant & Devine, 1998). Each of these two scales consists of five items to which participants responded on a 9-point bipolar scale (1 = strongly disagree; 9 = strongly agree). Finally, participants responded to several demographic questions.

Procedure

Upon arrival to the lab, participants were told that they would view images, answer questions about the images, and answer questions about themselves. After informed consent was obtained, participants were seated at a computer, were told that all instructions would be presented on the screen, and instructed to begin the experiment. The experiment was presented entirely on the computer. The experimenter was always a White female.

Participants in the White and Black prime conditions were first informed that they would view several images of faces, asked to view the faces, and told they would be asked about them later. While participants were not later asked about the faces, they were told this in order to assure that they would attend to the images. Participants in the control condition were not presented with any faces and were routed directly to the object detection task.

Participants then completed the object detection task, responded to items from the IMS and EMS, and completed demographic questions.

Results

For each object, we recorded the image number in the set on which the participant responded with the object name: we defined this as the step number. Lower step numbers indicate a faster response. Trials on which participants provided an incorrect response or did not respond were excluded from the analyses. This left us with a total of 411 object responses. The step on which participants correctly responded ranged from 3 to 38 with a mean of 19.06 (sd = 6.94).

We submitted the step data to a 3 (race of prime) x 2 (object type) mixed-model analysis of variance (ANOVA). The analysis revealed a non-significant but marginal main effect of race of prime on step number, $F(2, 408) = 2.79$, $p = 0.0628$), and a significant main effect of object type on step number, where weapon objects were identified on a significantly lower step number ($M = 17.43$) than were non-weapon objects ($M = 19.72$), $F(1,409) = 9.33$, $p = .0024$. For the analysis of interest, however, there was not a significant interaction between race of prime and object type on step number, $F(2,405) = 1.69$, $p = .186$. See figure a.

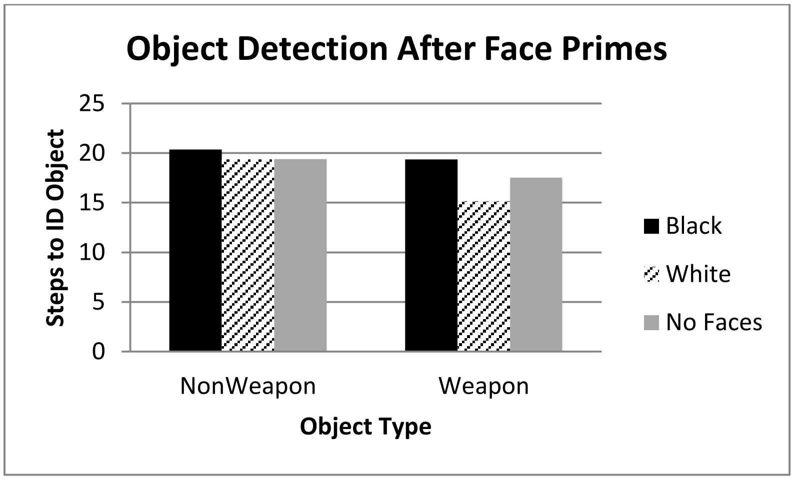
Discussion

In Pilot Study A, we did not find support for our prediction that viewing Black males faces would lead to faster detection of weapons in degraded images. There are several reasons we may have failed to replicate previous findings in this study. For one, because the faces were presented at a supraliminal level, participants may have noted the Black male faces and engaged in tactics to counter the automatic activation of stereotypes (see: Monteith, Mark, & Ashburn-Nardo, 2010). Participants may have also been more cautious when responding to the weapon objects after viewing the black male faces in order to avoid appearing prejudiced. While we presented the faces supraliminally in order to match the eventual supraliminal presentation of voices, this may have led to unintended effects.

The images we used may have had qualities unbeknownst to us that led to an elimination of the known effect of automatic activation of stereotypes. For example, the images of White males may have been particularly threatening, or the Black males in the images may have appeared particularly trustworthy. In future experiments, the images will be more carefully selected to match on dimensions that may have an effect on the phenomenon of interest.

However, the effect of race on weapon detection has already been established, and Pilot Study A was meant to act as a conceptual replication of that effect. With Pilot Study B, we hoped to extend this effect to our factor of interest: African American Vernacular English (AAVE).

Figure a. Number of steps to detect objects after face primes.



Pilot Study B

In Pilot Study B we attempted to extend Eberhardt and colleagues' (2004) study by examining whether listening to AAVE leads to faster detection of weapons. We predict that—consistent with the stereotypes elicited by visual facial race primes—listening to AAVE will lead to a faster detection of weapons in degraded images than listening to SAE.

Method

Participants

Forty-five White males at a downtown Chicago lab completed the study for a $5 payment. All participants signed up for an experiment described as a study of audio and visual perception. Six subjects were excluded from the analyses: 1 because he had already participated in Pilot Study 1 in another lab and 5 because of anomalies during the procedure (e.g., the participant was unclear on directions, the computer did not function properly, or the participant refused to respond to questions related to race). This left us with 39 participants whose data were included in the analyses. Participants were randomly assigned to a 3 (dialect of prime: Standard American English (SAE) prime (n = 14), African American Vernacular English (AAVE) prime (n = 15), or no-prime control (n = 10)) x 2 (object type: weapon or non-weapon) mixed-model design with dialect of prime serving as the between-subject factor and object type serving as the within-subject factor.

Materials

The detected objects used were the same as in Pilot Study A. The voices primes consisted of either White males speaking SAE or Black males speaking AAVE. The voice samples that participants heard consisted of 12 unique men stating, "Hi, my name is (given name, surname)."

The names that were used were generated to sound race neutral and were the same across conditions. The given names used were from a list of crossover names—names equally likely to be from White and Black males—generated using California birth certificate information from 1961 to the year available at the time of the list's creation (Levitt & Dubner, 2005). The surnames used were chosen by selecting surnames that were within 5 points in their rankings of commonness for self-identified white and black people on the United States Census in 2000 (Butler, 2013). For example, 'Stewart' is the 46^{th} most common surname among whites and 45^{th} most common surname among blacks in the United States, and was selected on the basis of therefore being common among both blacks and whites. Ranking was used rather than the raw number of whites and blacks with the surname because whites in the United States far outnumber blacks; therefore, using raw numbers would have led to the selection of extremely uncommon surnames or surnames that are somewhat common for blacks but uncommon for whites. That is, ranking data allows us to control for the baseline numbers of whites and blacks in the United States.

The 12 given names and surnames were then combined to produce 12 name pairs. While the names were paired somewhat arbitrarily, we switched any pairs that produced the full name of a reasonably famous person that participants may have recognized in order to eliminate distractions or confusion. We were left with 12 name pairs: George Stewart, Eric Evans, Dominic Parker, Cameron Taylor, Vincent Moore, Brandon Wilson, Shawn Davis, Frank Jones, Marc Williams, Gregory Johnson, Isaac Wright, and Craig Reed. We then recorded 12 white males who speak Standard American English and 12 black males who speak African American Vernacular English stating, "Hi, my name is (given name, surname)," using each of the 12 name pairs. The SAE speakers were White men affiliated with the psychology department, either as students or friends of students; the AAVE speakers were Black men who provided the recordings for payment at a downtown Chicago lab.

Procedure

Participants in the study were guided through the same procedure as in Pilot Study A, with the only exception being that rather than viewing Black or White faces during the prime block, they heard either White males speaking SAE or Black males speaking AAVE. The voice samples that participants heard consisted of 12 unique men stating, "Hi, my name is (given name, surname)."

Each participant only heard one dialect set, and the order of recordings within the set was randomized without replacement. Just as in Pilot Study 1, participants then completed the visual search task, responded to items from the Internal and External Motivation to Respond Without Prejudice Scales (IMS and EMS; Plant & Devine, 1998), and responded to demographic questions.

Results

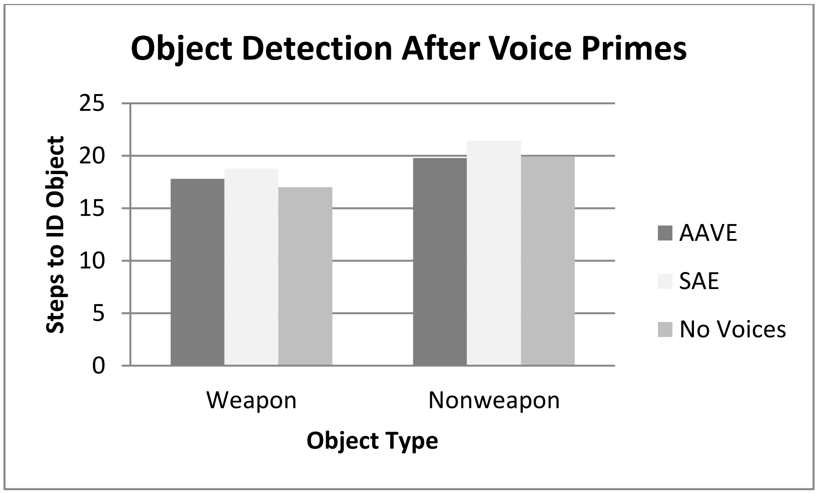
Just as in Pilot Study A, for each object, we recorded the image number in the set on which the participant responded with the object name: we defined this as the step number. Again, lower step numbers indicated a faster response. Trials on which participants provided an incorrect response or did not respond were excluded from the analyses. This left us with a total of 512 object responses. The step on which participants correctly responded ranged from 2 to 39 with a mean of 19.71 (sd = 7.14).

We submitted the step data to a 3 (dialect of prime) x 2 (object type) mixed-model analysis of variance (ANOVA). The analysis revealed a non-significant but marginal main effect of dialect of prime on step number ($F(2, 509) = 2.66$, $p = .0706$), and a significant main effect of object type on step number, where weapon objects were identified on a significantly lower step number ($M = 17.97$) than were non-weapon objects ($M = 20.4$), $F(1,510) = 12.47$, $p = .0005$. For the analysis of interest, however, there was not a significant interaction between dialect of prime and object type on step number, $F(2,506) = 0.16$, $p = .8503$. See Figure b.

Discussion

In Pilot Study B, we did not find support for our prediction that hearing African American Vernacular English would lead to faster detection of objects in degraded images. However, we were also unable to replicate Eberhardt and colleagues' (2004) findings in Pilot Study A using the same paradigm. Therefore, it is difficult to interpret these null findings. In order to determine whether AAVE elicits threat-related stereotypes, we first needed to find a paradigm that elicits known effects in our sample population. In Pilot Study C, we shifted our study design to incorporate a new paradigm.

Figure b. Number of steps to detect objects after voice primes.



Pilot Study C

Consistent with Eberhardt and colleagues' (2004) findings that priming Black male faces leads to faster detection of weapons, Payne (2001) also found that priming Black male faces leads to faster categorization of weapons (Experiment 1) and greater misidentification of tools as weapons when participants are prompted to respond quickly (Experiment 2). In Pilot Study C, we sought to replicate these findings.

Method

Participants

Participants were recruited for this study on Amazon Mechanical Turk (mturk) for monetary compensation. There were no restrictions placed on gender or race in order to encourage participants to report their demographics accurately; the only requirement was that participants must be located in the United States. Our study was accessed 1,019 times. We ran the study with an intended cut-off point of 1,000 participants on mturk, but the study was accessed an additional 19 times by people who did not indicate that they had completed it on mturk. This occurs if a participant signs up for the study but does not indicate that they have completed the study within the allotted time, or 'returns' the study after deciding not to complete it, but have already begun the procedures. We were left with 1,019 participants. 290 participants did not complete the study in its entirety and therefore could not be included in the analyses. Of the remaining participants, 67 participants were excluded from the analyses because their IP addresses or user identifications were the same as another participant's, and the veracity of the account could not be verified. An additional 8 participants were excluded from the analyses because they accessed the study from outside of the United States. In total, 654 participants' data were included in the statistical analyses.

Materials

During this study, participants completed the Weapons Identification Task (WIT; Payne, 2001). During the WIT, participants are told that they will first view a picture of a face, and then either a weapon or a tool. Participants are told to ignore the face, which simply serves as a warning that the weapon or tool is about to appear. They are told to determine whether the second image is of a weapon or a tool, and to press the 'E' key on the keyboard if the image is of a weapon, or the 'I' key on the keyboard if the image is of a tool. During this task, participants first viewed either a Black male face, White male face, or a neutral grey image for 200 ms, then viewed either a weapon or tool for 200 ms. An image mask (a black and white image of visual noise) then appeared until participants pressed either the 'E' or 'I' key on the keyboard.

The face primes and target images were all obtained from Keith Payne's laboratory web page (Payne, n.d.). All of the images of Black and White males were cropped to rectangles that only include the face of the male. The guns and tools are surrounded by a white, rectangular background.

The prime consisted of one of either 12 Black male faces or 12 White male faces, or a neutral grey box, that were presented at random. The target consisted of one of either 6 guns or 6 tools, which were also presented at random. Therefore, each trial consisted of a face or neutral prime and a target. Participants completed 12 practice trials, which were not scored, then 192 test trials. Reaction times (RTs) were recorded, as well as whether the response was correct or incorrect.

Procedure

Participants signed up for the task on Amazon mechanical turk (mturk). At this point, participants were only told that they would complete a study on categorization and perceptions. Once they accepted the study on mturk, participants were directed to click on a link to a Qualtrics survey. Once at Qualtrics, participants were presented with and accepted an informed consent form and provided their mturk ID numbers. They were then directed to a link on Inquisit, where they could download the Inquisit software to complete the WIT. By using Inquisit, we were able to record accurate reaction times from an online sample. After completing the WIT on Inquisit, participants were redirected to Qualtrics to complete demographic questions and be debriefed.

Results

Each of the 654 participants included in the analyses completed 192 scored trials, totaling 125,568 trials. In order to analyze reaction times, we excluded any trials on which participants responded faster than 100 ms or slower than 1,000 ms. This was also done a priori in Payne's (2001) original study according to standards laid out in Fazio (1990). There were 6,822 trials outside these boundaries. Further, we only included the trials on which participants correctly categorized the weapons and tools; there were 5,240 additional incorrect trials dropped. This left us with 113,506 trials included in our analyses of RTs.

All statistical analyses (unless otherwise noted) were performed using linear mixed effects models (Baayen, Davidson, & Bates, 2008) in R (R Development Core Team, 2016) and used the R packages lme4 (Bates, Maechler, Bolker, & Walker, 2015).

The main mixed-effects models of the results included the following fixed effects: Face Prime: White male face, Black male face, or a neutral grey square and Target Type: Tool or Weapon. Random effects were included where appropriate and consisted of random intercepts for subject and for prime item (i.e., specific prime file used). The dependent variable for each of these models was participants' log-transformed RT, or the time it took in milliseconds for participants to categorize the tool or weapon image once it was presented.

We tested whether the interaction between Face Prime and Target Type influenced participant RT. The method of analyses described above compares a model with this interaction term to one without and conducts a Chi-square goodness of fit test. From this, we ascertained that Face Prime and Target Type interacted to predict RTs ($\chi^2(2) = 36.82$, $p < .001$, $\varphi = 0.24$). We conducted a follow-up analysis comparing the Face Prime by Target Type interaction in only those trials on which the Face Prime was Black or White, thereby excluding the control trials. This 2 x 2 analysis was also significant ($\chi^2(1) = 20.29$, $p < .001$, $\varphi = 0.18$).

In Payne's 2001 study, only non-Black participants were included, and other research in this area has only included White participants. To determine whether our effects were characteristic of only White participants or could be generalized to participants of all races and ethnicities, we ran our analyses again examining only White, non-Hispanic participants' data (n = 490), then again examining only non-White participants' data (n = 164). Among White participants, we found a significant interaction in the 3 x 2 analysis ($\chi^2(2) = 24.41$, $p < .001$, φ = 0.22), and in the 2 x 2 analysis involving only Black and White Face Primes ($\chi^2(1) = 15.83$, $p < .001$, $\varphi = 0.18$). Among non-White participants, we found a significant interaction in the 3x2 analysis ($\chi^2(2) = 13.40$, $p = .0012$, $\varphi = 0.29$), and in the 2 x 2 analysis ($\chi^2(1) = 4.52$, $p = .0335$, φ = 0.17).

Discussion

In Pilot Study C, we replicated previous findings that racial primes influence reaction time when categorizing tools and weapons: participants were faster at categorizing weapons after viewing Black male faces than after viewing White male faces and were faster at categorizing tools after viewing White male faces than after viewing Black male faces. This is consistent with previous research and informs us that this effect can be replicated with our current procedure and with our current participant pool.

However, we realized that our experiment inherited a confound from the initial WIT paradigm (Payne, 2001): the side of the screen and hand with which weapons and tools were categorized were not counterbalanced (see General Discussion of main text for further discussion of this confound). For this reason alone, we re-ran this study to include left–right counterbalancing, as described in Study 1 of the main text.

Pilot Study D

In Pilot Study D, we sought to extend the effect of face primes on the automatic activation of stereotypes to determine whether voice primes also elicit crime-relevant racial stereotypes.

Method

Participants

Participants were recruited for this study on Amazon Mechanical Turk (mturk) for monetary compensation. There were no restrictions placed on gender or race in order to encourage participants to report their demographics accurately; the only requirement was that participants must be located in the United States. Our study was accessed 1,020 times. We ran the study with an intended cut-off point of 1,000 participants on mturk, but the study was accessed an additional 8 times by people who did not indicate that they had completed it on mturk, which left us with 1,008 participants. 322 participants did not complete the study in its entirety and therefore could not be included in the analyses. Of the remaining participants, 56 participants were excluded from the analyses because their IP addresses or user identifications were the same as another participant's, and the veracity of the account could not be verified. An additional 21 participants were excluded from the analyses because they accessed the study from outside of the United States. Because of the dual audio and visual nature of this study, participants were required to complete a manipulation check in which they were asked to indicate whether they recall any of the names stated by the voices; participants who did not list at least one correct name were excluded from the analyses. To this end, 114 participants were excluded from the analyses. While this points to a high manipulation check failure rate, it can be explained by the fact that many participants indicated they had technical audio difficulties and could not hear the sounds in the study; this may have had something to do with the participants' audio setup, but is highly unlikely to correlate with any predictive measures that would influence participant performance on this task. An additional 2 participants were excluded from the analyses because they did not follow directions and pressed the same response key for all items. In total, 493 participants' data were included in the statistical analyses.

Materials

During this study, participants completed a modified version of the WIT (Payne, 2001). During the modified WIT, participants were told that they will first hear a sound clip of a voice, and then see a picture of either a weapon or a tool. Participants are told to ignore the voice, which simply serves as a warning that the weapon or tool is about to appear. They are told to determine whether the second image is of a weapon or a tool, and to press the 'E' key on the keyboard if the image is of a weapon, or the 'I' key on the keyboard if the image is of a tool. During this task, participants first heard either a black male voice stating a name in AAVE, a white male voice stating a name in SAE, or silence, then viewed either a weapon or tool for 200 ms. An image mask (a black and white image of visual noise) then appeared until participants pressed either the 'E' or 'I' key on the keyboard.

The target images were all obtained from Keith Payne's laboratory web page (Payne, n.d.). The guns and tools are surrounded by a white, rectangular background. The prime consisted of one of either 12 black male voices speaking AAVE or 12 white male voices speaking SAE, or an interval of no sound as a control condition that lasted the same duration as the longest of the voice clips, that were presented at random.

The voice samples that participants heard consisted of 24 unique men—12 white and 12 black—stating a given and surname. The names that were used were generated to sound race neutral and were the same across conditions. The given names used were from a list of crossover names—names equally likely to be from white and black males—generated using California birth-certificate information from 1961 to the year available at the time of the list's creation

(Levitt & Dubner, 2005). The surnames used were chosen by selecting surnames that were within 5 points in their rankings of commonness for self-identified white and black people on the United States Census in 2000 (Butler, 2013). For example, 'Stewart' is the 46th most common surname among Whites and 45th most common surname among Blacks in the United States, and was selected on the basis of therefore being common among both Blacks and Whites. Ranking was used rather than the raw number of Whites and Blacks with the surname because Whites in the United States far outnumber Blacks; therefore, using raw numbers would have led to the selection of extremely uncommon surnames or surnames that are somewhat common for Blacks but uncommon for Whites. That is, ranking data allows us to control for the baseline numbers of Whites and Blacks in the United States.

The 12 given names and surnames were then combined to produce 12 name pairs. While the names were paired pseudo-randomly, we switched any pairs that produced the full name of a reasonably famous person that participants may have recognized in order to eliminate distractions or confusion. We were left with 12 name pairs: George Stewart, Eric Evans, Dominic Parker, Cameron Taylor, Vincent Moore, Brandon Wilson, Shawn Davis, Frank Jones, Marc Williams, Gregory Johnson, Isaac Wright, and Craig Reed. We then recorded 12 white males who speak Standard American English and 12 black males who speak African American Vernacular English stating each of the 12 name pairs. The SAE speakers were white men affiliated with the psychology department, either as students or friends of students; the AAVE speakers were black men who provided the recordings for payment at a downtown Chicago lab.

The target consisted of one of either 6 guns or 6 tools, which were also presented at random. Therefore, each trial consisted of a voice or neutral prime and a target. Participants completed 12 practice trials, which were not scored, then 192 test trials. Reaction times (RTs) were recorded, as well as whether the response was correct or incorrect.

Procedure

Participants signed up for the task on mturk. At this point, participants were only told that they would complete a study on categorization and perceptions. Once they accepted the study on mturk, participants were directed to click on a link to a Qualtrics survey. Once at Qualtrics, participants were presented with and accepted an informed consent form and provided their mturk ID numbers. They were then directed to a link on Inquisit, where they could download the Inquisit software to complete the modified WIT. Following the modified WIT, participants we redirected to Qualtrics to complete demographic questions and be debriefed.

Results

Each of the 493 participants included in the analyses completed 192 scored trials, totaling 94,656 trials. In order to analyze reaction times, we excluded any trials on which participants responded faster than 100 ms or slower than 1,000 ms. This was also done a priori in Payne's (2001) original study according to standards laid out by Fazio (1990). There were 3,560 trials outside these boundaries. Further, we only included the trials on which participants correctly categorized the weapons and tools; there were an additional 3,638 incorrect trials dropped, leaving us with 87,458 trials included in our analyses of reaction times.

All statistical analyses (unless otherwise noted) were performed using linear mixed effects models (Baayen, Davidson, & Bates, 2008) in R (R Development Core Team, 2016) and used the R packages lme4 (Bates, Maechler, Bolker, & Walker, 2015).

The main mixed-effects models of the results included the following fixed effects: Voice Prime: Standard American English (SAE), African American Vernacular English (AAVE), or time-matched silence and Target Type: Tool or Weapon. Random effects were included where appropriate and consisted of random intercepts for subject and for specific prime file used. The dependent variable for each of these models was participants' log-transformed RT, or the time it took in milliseconds for participants to categorize the tool or weapon image once it was presented.

We tested whether the interaction between Voice Prime and Target Type influenced participant RT. The method of analyses described above compares a model with this interaction term to one without and conducts a Chi-square goodness of fit test. From this, we ascertained that Voice Prime and Target Type interacted to predict RTs ($\chi^2(2) = 32.75$, $p < .001$, $\varphi = 0.26$). We conducted a follow-up analysis comparing the Voice Prime by Target Type interaction in only those trials on which the Voice Prime was SAE or AAVE, thereby excluding the control trials. This 2 x 2 analysis was not significant ($\chi^2(1) = 0.96$, $p = .3275$, $\varphi = 0.04$).

In Payne's 2001 study, only non-Black participants were included, and other research in this area has only included White participants. To determine whether our effects were characteristic of only White participants or could be generalized to participants of all races and ethnicities, we ran our analyses again examining only White, non-Hispanic participants' data (n = 355), then again examining only non-White participants' data (n = 138). Among White participants, we found a marginal interaction in the 3 x 2 analysis ($\chi^2(2) = 29.30$, $p < .001$, $\varphi = 0.29$), but not in the 2 x 2 analysis involving only SAE and AAVE Voice Primes ($\chi^2(1) = 1.03$, $p = .3096$, $\varphi = 0.05$). Among non-White participants, we did not find a significant interaction in the 3 x 2 analysis ($\chi^2(2) = 4.75$, $p = .0931$, $\varphi = 0.19$), nor in the 2 x 2 analysis involving only AAVE and SAE Voice Primes ($\chi^2(1) = 0.07$, $p = .7962$, $\varphi = 0.02$).

Discussion

In Pilot Study D, we did find evidence that Voice Primes influenced categorization of objects. However, we did not find evidence that this was specifically due to a difference in the time to categorize weapons and tools after hearing AAVE or SAE voices. Therefore, Pilot Study D does not give us sufficient evidence to address our experimental question.

In critiquing our design, we realized that by having participants hear the same name stated by both a Black male in AAVE and a White male in SAE, we may have inadvertently watered down any stereotypic associations activated by hearing the dialect.

Furthermore, because the voice clips were of varying lengths, we had programmed the target images to appear after the duration of the longest clip, creating pauses up to 200 ms after hearing some of the voices. Because it is yet unknown how any pauses between prime and target can influence reaction times on the WIT, we do not know whether this affected the outcome of our study. However, this choice introduced an unintended difference in timing between the Face and Voice versions of our WIT task.

Finally, in Pilot Study D, the side of the screen and hand with which tools and weapons were categorized were not counterbalanced. This confound is discussed in the General Discussion of the main article, and was taken into consideration when redesigning the studies reported in the main text of the article.

In redesigning our studies as described in Studies 2a and 2b, we addressed each of these issues. Rather than present participants with AAVE and SAE voices samples that use the same names, we doubled our pool of names so that no participant would hear the same name stated in more than one dialect. We decided to insert the necessary pauses before the voice clips so that the target image would appear as soon as the voice clip had ended. In re-running this study as described in Studies 2a and 2b, we also included left–right counterbalancing to rule out alternative explanations for our results. We also re-ran Pilot Study C in order to include left–right counterbalancing, as described in Study 1 of the main text of this article.