

Info 290: Essay 2
AIR MAIL
University of California, Berkeley

Introduction

Though many aspects of our current court dynamics encourage just decisions (e.g., specialized training of judges and screening of juries for “objective” rulings), our system contains intrinsic biases due to the subjective nature of human beings. When race is made salient -e.g., a black defendant and a white victim-, it is more likely that the defendant will experience a harsher punishment if guilty (Eberhardt, Davies, Purdiu-Vaughns, & Johnson, 2006; Baumgartner, Grigg, & Mastro, 2015).

This is not necessarily the fault of the law, but an error in human judgments (Eberhardt et al., 2006; Baumgartner et al., 2015). Some of the underlying human mechanisms causing this injustice are cognitive heuristics. By our nature, humans have cognitive limitations and rely upon mental shortcuts (Simons & Chabris, 1999; Kahneman, Slovic, & Tversky, 1982; Eberhardt et al., 2006). We will specifically focus on two of the many heuristics that contribute to bias in our decision making: the availability and representative heuristics.

Today people are bombarded by the stereotypical caricatures of minorities. For instance, black and Hispanic individuals portrayed as villains alongside white protagonists. (Simons et al., 1999; Kahneman et al., 1982). Since those stereotypes are prominent in today’s society, judges and juries are, to some degree, implicitly subject to them. When race is made salient (e.g., black defendant and white victim), many judges and jurors believe the defendant to be guiltier of a particular crime than the defendant actually is because that mental scenario is easily recalled into their minds (availability heuristic) and the defendant properly—albeit erroneously--represents what a “criminal” may look like (representativeness heuristic) (Eberhardt et al., 2006). This racial bias has been previously dubbed the race-of-victim effect (U.S. General Accounting Office, 1990).

Motivational theories within psychology give rise to another possible explanation for the inequity in ruling. Social identity theory states, “a person’s self-concept and self-esteem derive from the status and accomplishments of the various groups to which the person belongs” (Tajfel & Turner, 1979). Consequently, derogating others in the outgroup improves one’s self-esteem (Fein & Spencer, 1997). Following Fein et al.’s (1997) theoretical framework: when race is made salient within a court case and the defendant or victim is of the judges’/jury’s outgroup, those with low self-esteem before the trial may be more inclined to put down the defendant of the outgroup with a harsher ruling.

The proposed study is made up of two studies to investigate if there is a difference in people’s presumption of innocence, verdict, and harshness of their ruling toward *avatars* instead of people. Study 1 is a replication of previous research with extensions. Study 1 will randomly assign participants to one of four treatment conditions where the stimuli are pictures of real people. Study 2 will be the same as Study 1 except having the “cartoonization” of the treatment conditions for avatar likeness. In other words, Study 1 will form a baseline measurement to compare the findings from Study 2.

Study 1

Study 1’s research investigates how each condition (please see *Appendix A*) affects the judgements and other psychometrics of white U.S. citizens in hopes of gaining a deeper understanding of how the issue of race plays out in the courtroom. The present study is different

from previous research in that it explicitly controls for various factors simultaneously, which are:

- Participants' race (White Americans since they are the most prominent in US juries and judges);
- A crime vignette, which will remain constant across all treatment conditions, while manipulating the race of the defendant and victim;
- Asking participants to rate how innocent they presume the defendant to be, their verdict, and the harshness of their sentence (see below for further explanation);
- Measuring participants implicit racial biases via an Implicit Association Test;
- And the use of other psychometrics (e.g., Openness to Others, etc.).

Methods

Participants

Stratified random sampling will be used through an online recruiting platform (Prolific Academic) for a total of 500 adult (18 +) white participants, all of whom reside in the United States. Since the majority of judges and jurors in the United States are white, this study will choose only white participants to reduce variance. All 500 participants will be used in the final dataset for analysis and will be paid \$3.25 for their participation (if funding permits). This study does not intend to use clustering since individual level random assignment will be used.

Research Design

This study will use Qualtrics. It will be a 2 (black or white defendant) x 2 (black or white victim) between subjects design, meaning each participant evaluates only one of the four conditions. Participants will be randomly assigned to one of the four conditions. There will be ~125 participants in each condition/treatment. (Please refer to *Appendix A* to view each of the four conditions/treatments).

To enhance believability in the study's realness, two distractor cases will be used in addition to the experimental condition to sway participants from thinking the study was about the issue of race. Cases will be presented in random order to control for order effects. The stimuli of the defendants and victims in the four experimental treatment conditions will be retrieved from the Chicago Face Database – a database of standardized facial images of various people for research purposes. Five independent raters will score a handful of the database's images on age, physical attraction, and emotional neutrality via a 1 – 5 Likert scale. We will use only those images that scored similar on those factors in the official study. Then, we will use Photoshop to place the experimental images in driver's licenses and masked the personal identifier sections of the driver's license with tape.

A crime vignette (*Appendix B*) will be simultaneously presented with the experimental stimuli for participants to evaluate; the crime vignette will be presented to participants as a juror's note from an actual case to enhance believability. In many criminal court trials, the judge will allow juror members to take notes on the case to help inform their decisions. Rectangular blocks will also be used to block out sections of the vignette that is believed to be personal identifiers. The crime vignette used for the experimental conditions will purposely invoke a sense of ambiguity in a 2nd degree murder case to help tease out the participants' implicit biases (ambiguity is common amongst criminal court trials, such as the notorious Treyvon Martin case). For control purposes, we will not include a motive for the crime, relationship to the

victim, the crime's heinousness, and other variables needed to decide their verdict in a real court trial. Including those variables would, without a doubt, influence a jury's or judge's decision (Oswald, Bieneck, & Hupfield-Heinemann, 2009). Previous literature suggests introducing one variable in a crime vignette as it puts participants under a focused cognitive overload (Oswald et al., 2009)—in this case, it is the macro level of ambiguity. This is purposely done since implicit biases thrive under ambiguity (Simons et al., 1999; Kahneman et al., 1982; Eberhardt et al., 2006).

Measures

The main outcome measures for this study are participants' presumption of innocence of the defendant, their verdict (guilty/not guilty), and the harshness of their verdict. Other secondary measures are intended to be used alongside other covariates, such as education level, political affiliation, etc. Blocking education level is ideal since those who are educated may be more knowledgeable of racial injustices and their underlying mechanisms.

Presumption of Innocence. How innocent do you presume the defendant to be? (Likert scale: 1 – Not innocent at all to 5 – Extremely innocent)

Verdict. Please choose a verdict: Do you believe the defendant to be innocent or guilty? (Binary choices; innocent or guilty)

- If a guilty verdict is chosen, the following question will be asked on the next page:
 - *From the following, choose a sentence for the defendant (minimum for 2nd degree murder is 15 years):* (Five answer choices will be presented in 15-year increments with *Plea bargain of <15 years in prison* being the minimum and *Life in prison without the possibility of parole* being the maximum sentence.)

Race-Weapons Implicit Association Test (RW-IAT). The RW-IAT will be used to measure participants' implicit biases for mapping out relationships to each of the Independent Variables (Greenwald et al., 1998). Though the weapons in the RW-IAT are not modern-day weapons (see *Appendix C*), Brian Nosek—the creator and person who empirically established the RW-IAT—discusses, civil war weapons were chosen because of wanting as little plausibility of frequent observation with Black people compared to White people, measuring macro, abstract implicit racial associations (2017). The RW-IAT will be used in this study to operationalize people's implicit biases in a pairing task between race and weapons (since the controlled vignette in this study purposely elicits a scenario in which the defendant is being charged of using a weapon against the victim). Measuring participants' implicit biases is key to understand which condition their presumption of innocence, verdict, and harshness of their sentence are more just.

The following measures are intended to track potential covariates.

Social Dominance Orientation scale (SDO). The SDO scale will be used to measure one's preference for social inequality (Ho et al., 2015). People who score high on SDO, believing certain groups should dominate others, are more likely to derogate members of an outgroup who are, in their eyes, inferior (Ekehammar, Akrami, Gylje, & Zakrisson, 2004; Ho et al., 2015). This scale can inform our research question by mapping out how those who score high on SDO tie

into their judgements of how innocent they presume the defendant to be, their verdict, and harshness of their sentence across each of the treatment conditions..

Agreeableness and Open-Mindedness from the 2nd revision of the Big Five Inventory (BFI-2). The personality domains of the BFI-2 will also be used (Soto & John, 2017). When compared to the Big Five Inventory (Soto & John, 2009), the BFI-2 has been empirically established as having a robust hierarchical structure, controlling acquiescent responses with more balanced true- and false-keyed items on both domain and facet level (Soto et al., 2017). Agreeableness is used to measure how warm, friendly, etc. a person may be, while Open-Mindedness measures one's intellectual curiosity, innovativeness vs. consistency, and/or cautiousness. The literature is slim on empirically established personality domains as it relates to people's prejudices, but some found the Agreeableness and Openness domains (used from the 1st version of the Big Five Inventory) to be significantly related toward other scales measuring prejudice, specifically SDO and Right-Wing Authoritarianism (RWA) (Ekehammar et al., 2004).

Openness to Other Scale (OtO). The *Openness to Other* scale was recently constructed and validated to measure one's appreciation of other people different from their self (Antonoplis & John, 2017). This scale is different from the BFI-2's domain of Agreeableness in that Agreeableness measures one's temperament toward others—not their appreciation for culturally different people (Antonoplis & John, 2017; Soto & John, 2017). Since this is a novel scale, finding significance may provide a new theoretical lens in which to understand how participants presume the defendants to be, choose their verdict, and harshly sentence the defendants across each of the Independent Variables.

Study 2

Study 2's purpose is to understand if there is a difference in people's presumption of innocence, verdict, and harshness of their ruling toward *avatars* instead of people. And if so, what is the difference? The theoretical concept supporting Study 2 is the Media Equation (Reeves & Nass, 1996). Reeves et al. have empirically established via an accumulation of studies that people tend to treat computers and media as if they are real people and things (1996). Through this theoretical framework, findings from Study 1 should most likely be replicated by Study 2. Study 2 aims to further understand if virtual reality can be implemented in real court trials for more just rulings. This is investigated by masking the racial characteristics of a defendant and/or victim with other racial characteristics to hinder the onset of implicit racial biases (e.g., black victim with a white avatar to give them more human likeness, etc.). Doing so would, theoretically, take away the defendant's white privilege of being seen as more innocent, leading to more just rulings overall. Ideally, avatars may be used in court trials to mask the racial characteristics of the defendant and/or victim in real-time to create more just rulings.

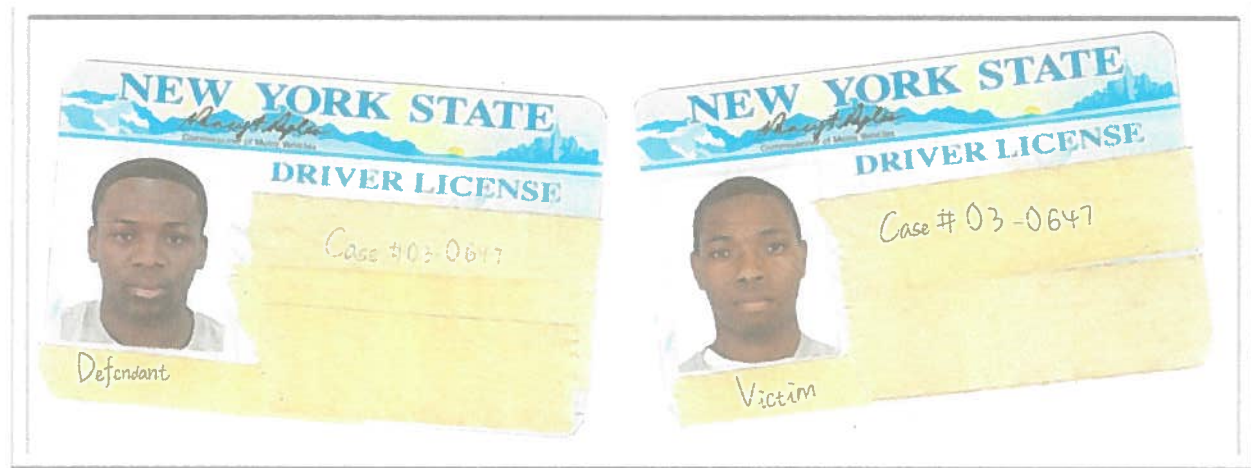
The only difference between Study 1 and Study 2 is Study 2 will use the cartoonization of each of the treatment conditions to represent "avatar likeness". Everything else (sampling methodologies, measurements, etc.) from Study 1 will be used in Study 2.

Appendix A:

	White Victim	Black Victim
Black Defendant	Treatment 1	Treatment 2
White Defendant	Treatment 3	Treatment 4



Treatment 1



Treatment 2



Treatment 3



Treatment 4

Appendix B:

[REDACTED]

[REDACTED]

[REDACTED]

Victim was discovered in Central Park by a couple who were jogging through. The couple found Defendant standing over Victim body with blood on his hands.

Victim had blunt force trauma to the head and was unconscious. After the couple called 911, the police arrived and took Defendant statement. Defendant claims to have seen Victim bleeding face down on the ground and says he only attended to help him.

The police ran Defendant ID and discovered that he has a criminal record for misdemeanor assault.

Victim died on the way to the hospital. Post-mortem examination provided evidence that the cause of death was due to blunt force trauma to the head. Now the prosecutor ~~believes~~ believes Defendant committed the crime and should be charged with Victim death as 2nd degree murder. The defendant claims to be innocent.

Figure 5: Crime Vignette (presented to participants as a Juror's note).

Crime vignette text:

Victim was discovered in Central Park by a couple who were jogging through. The couple found Defendant standing over Victim body with blood on his hands. Victim had blunt force trauma to the head and was unconscious. After the couple called 911, the police arrived and took Defendant statement. Defendant claims to have seen Victim bleeding face down on the ground and says he only attended to help. The police ran Defendant ID and discovered that he has a criminal record for misdemeanor assault. Victim died on the way to the hospital. Post-mortem examination provided evidence that the cause of death was due to blunt force trauma to the head. Now, the prosecutor believes Defendant committed the crime and should be charged with Victim death as 2nd degree murder. The defendant argues to be innocent.

Appendix C:

**African
American
faces**



**European
American
faces**



weapons



**harmless
objects**



References

- Antonoplis, S. & John, O. (2017). Openness to Other Construction and Validation. Manuscript in preparation.
- Baumgartner, F. R., Grigg, A. J., & Mastro, A. (2015). # BlackLivesDon'tMatter: race-of-victim effects in US executions, 1976–2013. *Politics, Groups, and Identities*, 3(2), 209-221.
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological science*, 17(5), 383-386.
- Ekehammar, B., Akrami, N., Gylje, M., & Zakrisson, I. (2004). What matters most to prejudice: Big Five personality, Social Dominance Orientation, or Right-Wing Authoritarianism? *European Journal of Personality*, 18(6), 463-482.
- Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of personality and Social Psychology*, 73(1), 31.
- Greenwald, Anthony G., McGhee, Debbie E., Schwartz, Jordan L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480.
- Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO7 scale. *Journal of Personality and Social Psychology*, 109(6), 1003-1028.
- Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M. Ranganath, K. A. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18(1), 36-88.
- Oswald, M. E., Bieneck, S., & Hupfield-Heinemann, J. (2009). *Social Psychology of Punishment of Crime*. Malden, MA: John Wiley & Sons, Ltd. Publications.
- Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, New York: CSLI Publication, Center for the Study of Language and Information.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059-1074.
- Soto, C. J. & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, 43, 84-90.
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117-143.
- Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56-65.