# Signaling Race, Ethnicity, and Gender with Names: Challenges and Recommendations*

Elizabeth Mitchell Elder[†] Matthew Hayes [‡]

**Abstract**

A growing body of research uses names to cue experimental subjects about race, ethnicity, and gender. However, researchers have not explored the myriad of characteristics that might be signaled by these names. In this paper, we introduce a large, publicly available database of the attributes associated with common American first and last names. For 1,000 first names and 21 last names, we provide ratings of perceived race; for 336 first names, we provide ratings on 26 social and personal characteristics. We show that the traits associated with first names vary widely, even among names associated with the same race and gender. Researchers using names to signal group memberships are thus likely cuing a number of other attributes as well. We demonstrate the importance of name selection by replicating DeSante (2013). We conclude by outlining two approaches researchers can use to choose names that successfully cue race (and gender) while minimizing potential confounds.

*The authors would like to thank Alexa Bankert, Diana Z. O'Brien, Leslie Schwindt-Bayer, Beth Simas, Bob Stein, Randy Stevenson, Rick Wilson, and numerous conference participants for valuable feedback and suggestions. We also thank Christopher DeSante for sharing his original experimental materials with us.

[†]Postdoctoral Researcher, Princeton University.

[‡]Assistant Professor, Department of Political Science, Rice University.

Discrimination on the basis of race, gender, and ethnicity remains a serious problem in American public life, yet measuring such discrimination has proven difficult due to social desirability bias. In recent years, scholars have turned to audit studies and survey experiments that use a minimal cue of race, ethnicity, or gender: individuals' names. Doing so has allowed scholars to shed light on potential biases by employers (Bertrand and Mullainathan, 2004), elected officials (Butler and Broockman, 2011), and ordinary citizens (DeSante, 2013) without the potential confound of social desirability bias.

Despite the prevalence with which names are used to indicate social group memberships, it is not clear that group memberships of interest are the only attributes triggered when people encounter particular names. Fryer Jr and Levitt (2004), for example, raise the concern that experiments using distinctive African American names may also be cuing subjects about individuals' class backgrounds, which could potentially compromise these studies' internal validity (but see Butler and Homola, 2017). Thus, studies showing racial bias might be detecting both taste-based discrimination against minority groups and bias on the basis of perceived traits of those group members, such as their class or educational status.[1]

In this study, we provide the most in-depth assessment of the groups, traits, and stereotypes that are invoked by the use of common American first names across different racial, ethnic, and gender categories. Using ratings from online respondents, we create two datasets. In our first dataset, we measure the perceived racial distinctiveness of 1,000 common American first names across the four largest racial and ethnic groups in the United States. In our second dataset, we select a subset of these names and measure not only how people perceive their racial distinctiveness, but also how they perceive those names on an additional 21 attributes such as warmth and competence. Using these two datasets, we find that the traits associated with first names vary widely, even within the same race and gender, suggesting that researchers using names to cue particular group memberships are likely cuing other attributes as well. We then demonstrate the importance of selecting a race or gender prime by replicating the findings of DeSante (2013) using purposefully-selected first names. Finally, we propose two methods with which researchers can choose names that cue race and gender while minimizing potential confounds.

## Information conveyed by names

Americans' names, both their given and surname, provide important clues to their identity and group membership. In the Black community, for example, unique naming conventions emerged as part of identity-building during the Civil Rights Movement (Lieberson and Mikelson, 1995), as well as earlier as responses to slavery, colonialism,

[1]It is important to note that such perceptions can be, in and of themselves, a manifestation of racial bias. Indeed, our data show that African American names are rated lower than white names on a number of desirable traits. This effect is especially pronounced for African American women's names. For more detailed analysis, see Section S3.2 in the Supplemental Information.

and emancipation (Cook, Parman and Logan, 2021). As a result, distinctively Black names, much like names with Spanish- or Asian-language origins, can convey useful information about an individual's racial/ethnic background to observers.

This information provides both an opportunity and a challenge for experimental researchers. On the one hand, first and last names can be used successfully to cue observers or experimental subjects about an alias's racial and ethnic identity (Butler and Homola, 2017; Gaddis, 2017*b*,*a*; Crabtree and Chykina, 2018). On the other hand, these names may combine a racial or ethnic signal with other information about the individual (Dafoe, Zhang and Caughey, 2018). In particular, scholars have pointed to names' abilities to provide cues about social class (Barlow and Lahey, 2018; Elchardus and Siongers, 2011), immigrant status (Gaddis, 2019*a*), age (Johfre, 2020), or other demographic characteristics. Additionally, name cues may activate not only demographic but also trait assessments. Two names equally likely to be assessed as African American may differ in key valence attributes such as warmth or competence (Newman et al., 2018).

To date, no study has provided a thorough exploration of the information conveyed by common American first and last names. This poses the risk that experimental researchers relying on name primes may be conflating the effects of racial or gender discrimination with inferences about an individual's personal traits. In our paper, we assess the information conveyed by a large set of common American first and last names. Our focus is on assessments of the typicality of a particular name among different racial/ethnic groups (African Americans, Asian Americans, Latinos, and Anglos), genders (men and women), class backgrounds (working, middle, and upper class), and the extent to which a name triggers various trait assessments or stereotypes.

## Data and methods

To assess what racial groups and traits are cued when individuals encounter names, we developed three publicly available datasets: the racial distinctiveness of 1,000 common American first names (our "Racial Distinctiveness" dataset), the groups and traits cued by a smaller subset of 336 first names (our "Name-Trait Perceptions" dataset), and the groups and traits cued by 21 common American first-last name pairs (our "Given and Surname Perceptions" dataset). To develop these datasets, we first obtained a list of given names of people born in the U.S. between the years 1955 and 1990 from the Social Security Administration (SSA). We merged the SSA data with data from Tzioumis (2018), which draws on proprietary mortgage loan data to provide the distribution of self-reported race for 4250 common first names. We identified the gender of first names using the most common gender of name recipients from the SSA data. We also obtained a list of common surnames by racial group from the Census Bureau data on the 1,000 most common surnames in the U.S. Further details on the selection of names to test can be found in Sections S2 and

S3 in the Supplemental Information.

We then fielded a series of surveys asking respondents to evaluate the groups and traits associated with these names. For each name, respondents were asked "How likely is it that someone named [NAME] is [GROUP/TRAIT]," where both the name and the group or trait being rated were randomly selected. Each trait or group rating was scored on a 5-point scale ranging from "Not likely at all" (1) to "Extremely likely" (5) with each response option labeled.

For our "Racial Distinctiveness" dataset, names were drawn from a list of 1,000 common first names, and groups were randomly drawn from one of four racial groups – African American, Asian American, Hispanic, and white. For our "Name-Trait Perceptions" dataset, names were randomly drawn from a smaller list of 336 first names, and groups/traits were randomly drawn from a list including: African American, Asian, Hispanic, white, a Democrat, a man, a Republican, a woman, aggressive, assertive, compassionate, competent, foreign, hardworking, honest, intelligent, likable, middle class, professional, strong, traditional, upper class, violent, warm, working class, and athletic. Finally, for our "Given and Surname Perceptions" dataset, first- and last-name pairs were randomly drawn from a list with 21 surnames and 126 first names, and groups/traits were randomly drawn from the same list as our "Name-Trait Perceptions" dataset.

To construct these datasets, we recruited respondents from Lucid Theorem from January-December 2020. Each respondent was asked to rate between 100 and 150 randomly selected names, although we allowed respondents to exit the survey before completing this number. We collected a total of 252,996 ratings from 4,792 respondents for our "Racial Distinctiveness" dataset, 371,189 ratings from 5,660 respondents for our "Name-Trait Perceptions" dataset, and 305,486 ratings from 2,660 respondens for our "Given and Surname Perceptions" dataset. After collecting these ratings, we estimated each name's score on each of the different attributes using a mixed effects model. We predict attribute ratings with fixed effects for name and random effects for each rater. We interpret the coefficient on a name's fixed effect (i.e. the intercept estimated for that name) as that name's level of the attribute in question. This approach allows us to estimate, for example, a name's perceived competence by combining all the ratings of that name's competence while accounting for individual raters' tendency to rate all names as competent. It also allows us to calculate standard errors for our estimates that take account of raters providing multiple ratings.

## Racial Associations of Common Names

We first present the results of the 1,000 name sample rated by Lucid Theorem respondents on four attributes: their likelihood of being held by someone who is white, Black, Hispanic, or Asian American. Figure 1 shows the rated likelihood of each name belonging to each race by the true likelihood of each name belonging to each race,

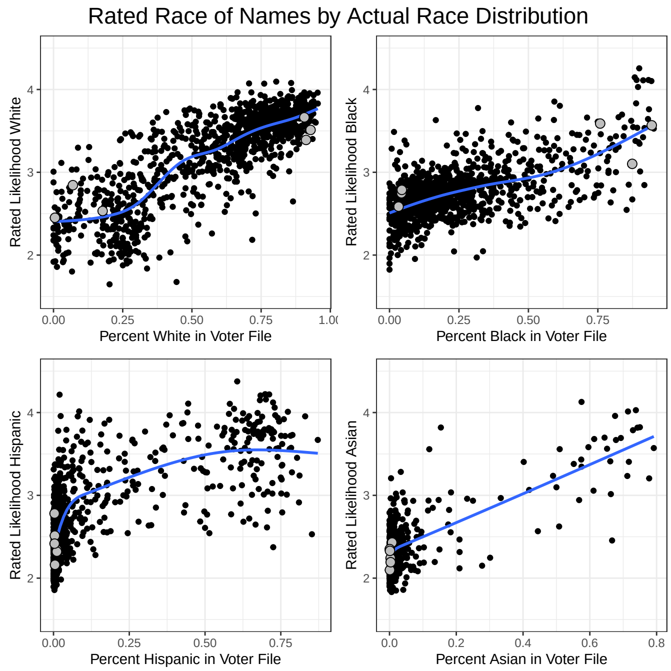


Figure 1: Raters accurately perceive the racial associations of common American names.

as measured by the percentage of nameholders belonging to that race in the North Carolina state voter file data.[2] The positive slopes in each of the panels of Figure 1 suggest that our raters, in general, perceive the race of names accurately: the larger the proportion of people with a name belong to a racial group, the likelier our raters were to say someone with that name belonged to that race.[3]

In a dataset available in the Supplemental Information, we assign each name a perceived race based on the racial group that received the highest likelihood rating for that name. We also assign each name a racial distinctiveness score based on the difference in rating between the highest and second-highest racial group. For example, the name Alberto has a Hispanic rating of 4.01 and a rating of 2.59 for its second-highest racial group, white. The name Alberto is therefore classified as Hispanic with a distinctiveness score of 1.42. This allows researchers to identify the names that most clearly cue membership in each racial group. We discuss recommendations for the use of this data in the Recommendations section below.

## Intra-group variation in ratings: Replication study

The data presented in the previous section can help researchers choose names that communicate membership in a racial group. However, names that are racially distinctive are often distinctive along other dimensions as well. For

[2]This data was obtained to complement the Tzioumis (2018) mortgage data; the North Carolina data contains better representation of African American names, and a larger sample of names overall.

[3]Figure 1 also highlights the perceived racial attributes of six names we investigate more closely below. The gray dots represent the names Latoya, Keisha, Octavia, Misty, Laurie, and Emily, moving left to right in the top-left panel.

example, Figure 2 shows the average ratings of names on the trait "competence" with 84% confidence intervals.[4] As you can see, even among names perceived to be white, there is substantial variation in perceived competence – Kathleen is perceived to be highly competent, whereas April is perceived to be low in competence. This underscores the importance of selecting names carefully to avoid distorting estimates of race/gender bias.

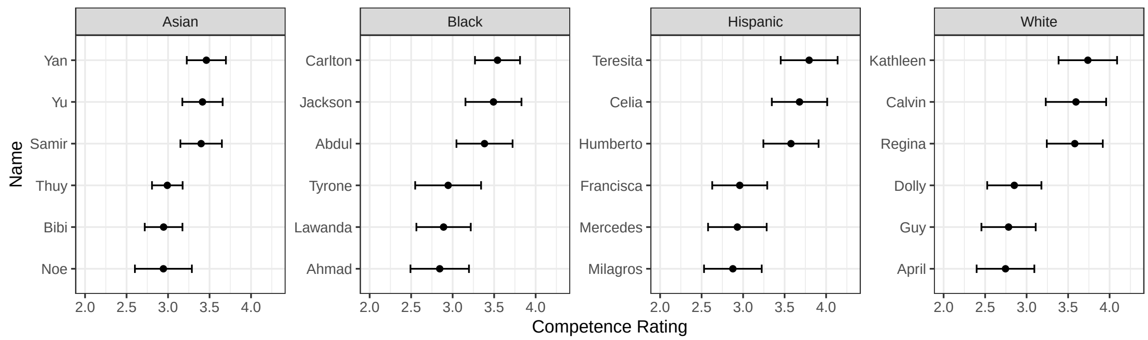


Figure 2: Competence ratings for selected names by perceived race

To further illustrate this point, we now partially replicate and extend the findings of DeSante (2013), a case where names that vary in nonracial traits produce substantively different results. In his article, DeSante (2013) presented respondents with pairs of hypothetical welfare applicants and asked them to allocate welfare funds to the applicants they viewed, generally finding evidence for racial discrimination. DeSante used the names Emily and Laurie to cue respondents to a white applicant, and Latoya and Keisha to cue respondents to a Black applicant. Figure 3 presents ratings of DeSante's names on several key dimensions.[5] From our data collection, it is clear that the names DeSante chose are good options for signaling race. Emily and Laurie are readily identified as white. Latoya and Keisha are similarly well-suited to signal that an individual is African American. These names, however, likely trigger additional traits and considerations.Figure 3 shows that Emily and Laurie are rated as more competent, hardworking, intelligent, professional, and warm than either Latoya and Keisha. While these differences in evaluation likely indicate racial bias in and of themselves, they also suggest a more complicated mechanism than simple racial phenotype for the racial discrimination found by DeSante (2013). In other words, it is possible that these traits are, in part, the mechanism through which racial bias is operating.

We can, however, select racially-distinct names that do not trigger these attributes. Figure 3 includes a racially distinct Black name (Octavia) and white name (Misty) that run counter to the trends in the names chosen by DeSante. Unlike Latoya and Keisha, Octavia is perceived as a competent ($\mu = 3.11$), hardworking ($\mu = 2.88$), and profes-

[4]A more thorough examination of intra-group variations in perceptions can be found in Section S3.1 of the Supplemental Information.

[5]These ratings were collected in a pilot study from a sample recruited via Amazon's Mechanical Turk. Full trait ratings can be found in Figure S11 in the Supplemental Information

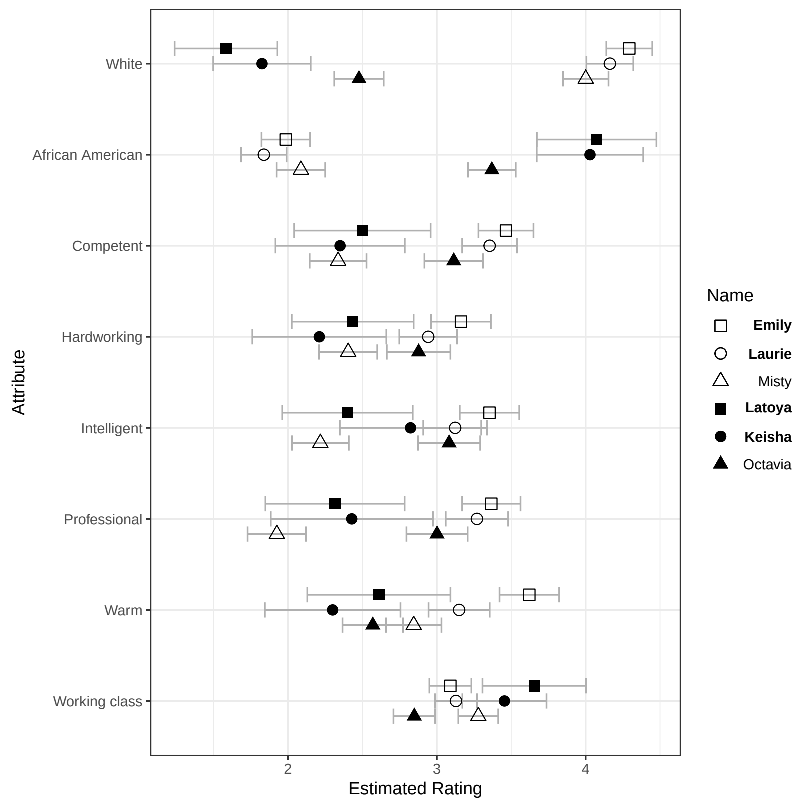


Figure 3: Average trait and group ratings by first name for selected women's names

sional ($\mu = 3.00$) person. Misty, in contrast, is perceived as less competent ($\mu = 2.34$), hardworking ($\mu = 2.40$), and professional ($\mu = 1.92$) than either Emily or Laurie. By selecting names purposefully to minimize these trait differentials, we can begin to identify the mechanisms through which racial bias operates (e.g., phenotype preference or taste-based discrimination vs. white respondents' biased perception that African Americans are less competent or hardworking). If part of the racial bias uncovered by DeSante (2013) is operating via inferences about worker competence, then we should see smaller treatment effects based on the traits accompanying racially-distinct names.

This leads us to hypothesize that both worker competence and worker race will affect the amount of welfare funding awarded by respondents. We expect that high-competence names (Emily and Octavia) should be awarded more funding than low-competence names (Misty and Keisha). While we expect a racial gap in funding to persist, we expect that the difference in funding between Black and white names should be smaller for a high-competence Black name (Octavia) than for a low-competence Black name (Keisha).

To test these hypotheses, we fielded an experiment modeled after the treatment used in DeSante (2013) on Amazon's Mechanical Turk in September, 2019. In the original study, all respondents saw two welfare applications with either no names, two white names (Laurie and Emily), a white and a Black name (Laurie and Keisha), or two Black names (Latoya and Keisha). These applications were also randomly assigned to have a “Worker Quality Assessment” of “Poor,” “Excellent,” or no assessment. Respondents were then asked to allocate money from a $1,500 budget to each application. Any money left unallocated would go to “offsetting the state budget deficit.”

In order to isolate the effects of race versus the traits people ascribe to racially-distinct names, our replication is a simplified version of this experiment. In our replication, respondents viewed two welfare applications identical in appearance to the original experiment. Rather than randomizing both applications, all respondents viewed the same baseline application of “Emily” who was rated as “Excellent” compared to a second application. We used a $2 \times 2 \times 2$ factorial design for this second application, randomizing the race (white/Black), competence (high/low), and quality assessment (excellent/poor) of the second applicant using the names Laurie, Misty, Octavia, and Keisha as our cues for race and competence.[6] Respondents were then asked to allocate funding to the two applicants or to offsetting the deficit. We collected a total 1,200 respondents across these 8 experimental conditions.[7]

Figure 4 presents the main results of our replication for treatments with a “Poor” worker quality assessment.[8] The first two panels are for white and Black names with more positive valence traits, respectively. What we find here is remarkable consistency; there are effectively no differences for Laurie or Octavia who are rated as poor workers. The average amount allocated to Emily in these conditions was $675 when paired with poor Laurie and $683 when paired with poor Octavia, and the average amounts received by Laurie and Octavia are nearly identical: $530.25 and $530.50. This suggests that racial bias is not uniform; Octavia and Laurie appear to be punished equally for having a poor assessment of their worker quality.

When respondents viewed excellent Emily paired with poor Keisha, however, they allocated significantly more ($745) to Emily than in any other condition ($p = 0.025$, $p = 0.008$, and $p = 0.021$ compared to Laurie, Misty, and Octavia, respectively). This provides some evidence of racial bias consistent with the original findings, although the form of that bias appears to be an increased reward for a white applicant rather than an increased penalty for a Black applicant. The only condition where we see some evidence of punishment is when the applicant is a poorly-rated Misty. In that condition, Misty received less money ($493) than other comparably-rated names ($p = 0.058$, $p = 0.081$, and $p = 0.084$ compared to Keisha, Laurie, and Octavia, respectively), with respondents allocating more to offset the state deficit.

Our findings suggest that the choice of names used to cue race is important, and can affect the type of bias we

[6] A discussion of how our design differs from DeSante (2013) can be found in Section S8.2 of the Supplemental Information.

[7] Pre-registration information with the Open Science Foundation can be found at `URL BLINDED FOR REVIEW`

[8] Results of “Excellent” worker quality assessments can be found in the Supplemental Information. They reveal minimal evidence for race or status differences.

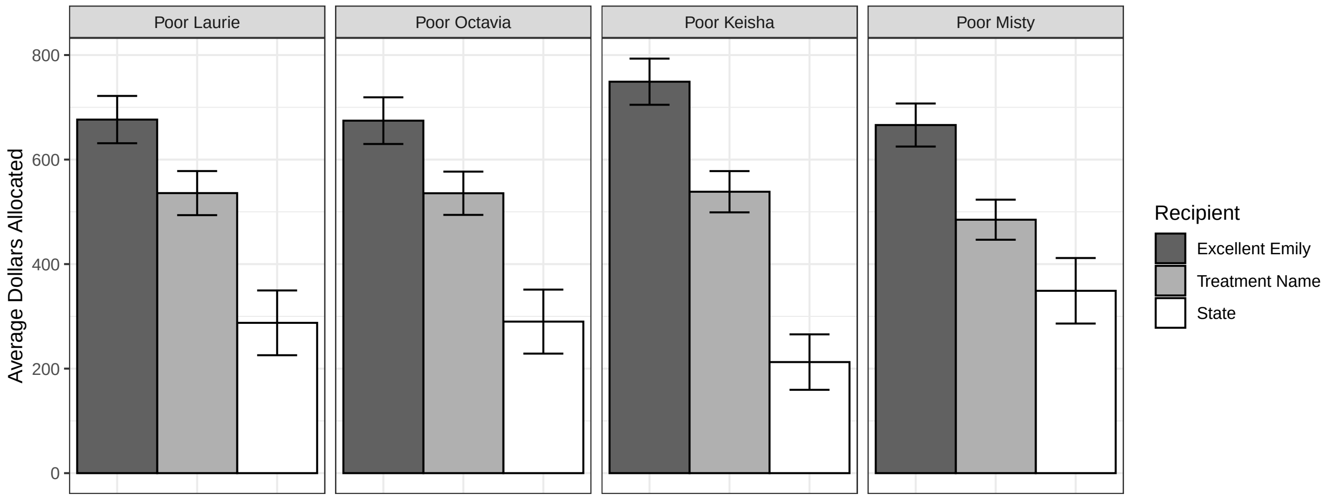


Figure 4: Average allocation by treatment condition

detect in our experiments. Using a high-status Black name, like Octavia, is likely to minimize racial differences. Indeed, if DeSante (2013) had used only white and Black names of comparable perceived status or competence, it is entirely possible that no evidence of racial bias would have emerged, as in our findings for Laurie versus Octavia. If lower status Black names are compared to higher status white names, we are likely to find larger treatment effects and a pro-white or anti-Black bias (in our Keisha condition). Finally, it is even possible to mistake class- or trait-based bias as anti-white bias if researchers select relatively lower status white names, such as Misty, compared to higher status Black names, such as Octavia.

## Recommendations for Future Research

In combination, our large, publicly available names datasets and empirical analyses yield important findings. On the one hand, we find that racially- (and gender-) distinct first and last names do send fairly clear signals about group membership. Combined with past studies (Butler and Homola, 2017; Gaddis, 2017*a*,*b*), this provides greater confidence in the use of such cues in experimental research. On the other hand, we demonstrate that names trigger a variety of attributes in addition to just race and gender. Building on previous studies that show names trigger other demographic traits, such as age (Johfre, 2020) or class (Gaddis, 2019*b*), we demonstrate that names also convey a wide range of personal traits and stereotypes. When individuals encounter a name such as "Emily," for example, they are likely inferring not just that she is a white woman, but also that she is relatively warm, competent, and intelligent. Importantly, we show that there are notable differences in these inferences across names within the same racial group.

What, then, should experimenters do? Rather than abandon the enterprise of measuring racial—and also gender—

effects using names, we argue that scholars should treat the complicated nature of such cues as a feature and not a bug. We recommend that experimenters interested in studying racial effects should take one of two approaches in selecting names, depending on their objectives: a matched approach or a randomization approach.

**The Matched Approach:** In some cases, researchers are interested in isolating the effect of perceived racial phenotype on evaluations, and there is clear theoretical and/or empirical guidance about other traits that might influence evaluations and correlate with race. Researchers in this situation should seek to minimize the differences between names used to cue different races, with special attention to differences on traits that they would theoretically expect to be associated with the outcome of interest. In this case, we suggest researchers use our "Name-Trait Perceptions" dataset to select names that are racially-distinctive, but similar on other characteristics that might affect evaluations. For example, audit studies exploring race-based job-market discrimination should select names that differ on race but are similar on traits desirable to employers, such as competence or laziness.

**The Randomization Approach:** Often, researchers are not interested in isolating the effects of race independent of other relevant attributes or there is insufficient guidance about what other traits may matter for evaluations. In this case, we suggest using our "Racial Distinctiveness" dataset to select a group of names that are distinctive for the race(s) of interest. Then, in implementing their experiment, researchers should randomize across these names (i.e., create multiple versions of the treatment containing different realizations of racially distinctive names). For example, a researcher creating a profile of a Black candidate in a survey experiment could use our list of the fifty most distinctively Black names, and display a randomly selected name for the candidate from this list to each respondent. This approach allows for a fuller representation of the universe of racially distinctive names and should prevent the idiosyncratic attributes of any particular name from unduly influencing the results. Using either of these approaches allows experimenters to measure racial bias while minimizing the impact of potential unanticipated confounds[9].

Beyond these two methods, there are many other ways researchers could incorporate our data into their design choices (see, for example, Acharya, Blackwell and Sen (2018)'s method for understanding causal mechanisms by selectively providing information on mediators). But more generally, our work shows that paying attention to the nature of treatments cuing race (or gender) can help us identify the scope of bias and uncover the mechanisms that drive it. DeSante's (2013) original findings, for example, suggest significant bias against Black women welfare applicants. We find that the results are more nuanced. Although racial bias clearly exists, it operates at least in part through stereotypes of African Americans as less competent and hardworking. And these stereotypes are not applied equally across all African American names. By separating out the potential pathways through which race, gender,

[9]See section 8.5 in the Supplemental Information for an application of this approach to the replication study reported above.

and ethnicity operate, we can make significant strides in understanding the nature of prejudice.

## References


Acharya, Avidit, Matthew Blackwell and Maya Sen. 2018. "Analyzing causal mechanisms in survey experiments." *Political Analysis* 26(4):357–378.

Barlow, M Rose and Joanna N Lahey. 2018. "What race is Lacey? Intersecting perceptions of racial minority status and social class." *Social Science Quarterly* 99(5):1680–1698.

Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal?" *The American Economic Review* 94(4):991–1013.

Butler, Daniel M and David E Broockman. 2011. "Do politicians racially discriminate against constituents? A field experiment on state legislators." *American Journal of Political Science* 55(3):463–477.

Butler, Daniel M and Jonathan Homola. 2017. "An empirical justification for the use of racially distinctive names to signal race in experiments." *Political Analysis* 25(1):122–130.

Cook, Lisa D, John Parman and Trevon Logan. 2021. "The antebellum roots of distinctively black names." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* pp. 1–11.

Crabtree, Charles and Volha Chykina. 2018. "Last name selection in audit studies." *Sociological Science* 5:21–28.

Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. "Information equivalence in survey experiments." *Political Analysis* 26(4):399–416.

DeSante, Christopher D. 2013. "Working twice as hard to get half as far: Race, work ethic, and America's deserving poor." *American Journal of Political Science* 57(2):342–356.

Elchardus, Mark and Jessy Siongers. 2011. "First names as collective identifiers: An empirical analysis of the social meanings of first names." *Cultural Sociology* 5(3):403–422.

Fryer Jr, Roland G and Steven D Levitt. 2004. "The causes and consequences of distinctively black names." *The Quarterly Journal of Economics* 119(3):767–805.

Gaddis, S Michael. 2017*a*. "How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies." *Sociological Science* 4:469–489.

Gaddis, S Michael. 2017*b*. “Racial/ethnic perceptions from Hispanic names: Selecting names to test for discrimination.” *Socius* 3:2378023117737193.

Gaddis, S Michael. 2019*a*. “Assessing Immigrant Generational Status from Names.” *Available at SSRN 3022217* .

Gaddis, S Michael. 2019*b*. “Signaling class: An experiment examining social class perceptions from names used in correspondence audit studies.” *Available at SSRN 3350739* .

Johfre, Sasha Shen. 2020. “What Age Is in a Name?” *Sociological Science* 7:367–390.

Lieberson, Stanley and Kelly S Mikelson. 1995. “Distinctive African American names: An experimental, historical, and linguistic analysis of innovation.” *American Sociological Review* pp. 928–946.

Newman, Leonard S, Mingxuan Tan, Tracy L Caldwell, Kimberley J Duff and E Samuel Winer. 2018. “Name norms: A guide to casting your next experiment.” *Personality and Social Psychology Bulletin* 44(10):1435–1448.

Tzioumis, Konstantinos. 2018. “Demographic aspects of first names.” *Scientific data* 5:180025.