

Flawed Measurement of Hiring Discrimination against African Americans

Raj Ghoshal¹

1. Elon University, Elon, NC, USA
rghoshal@elon.edu

Received July 26, 2019

Accepted for publication December 16, 2019

Published December 20, 2019

Abstract

This article addresses debates about race and hiring discrimination. Many experimental audit studies compare how equal-quality resumes with typically-white versus typically-black names are treated on the job market, often finding a difference in callback rates. However, one recent study (Darolia et al. 2016) argues that many employment audits may overstate discrimination because they use “black names” that are overly stereotypical. Pairing typically-white first names with typically-black last names to signal African American applicants’ race, they send out 9,000 resumes and find no discrimination in callbacks. Recognizing that some employers may not view the names they used, such as “Chloe Jackson,” as belonging to black applicants, they nonetheless argue that employers’ “error rate” in racial attribution would need to be “nearly 60 percent,” which they regard as improbably high, for prior audit findings to hold. Using a national survey of 1,050 Americans, I examine how people perceive the exact name combinations Darolia et al. used. I find that respondents’ overall error rate is indeed “nearly 60 percent” (59.7 percent), that this pattern holds for a subsample of respondents most likely to make hiring decisions, and that the pattern is unique to their approach rather than a problem for audit studies generally. I show that prior findings of persistent discrimination are not just unchallenged, but in fact bolstered by Darolia et al.’s results and consider implications for future work.

Keywords: Race, audit studies, employment, hiring, names, discrimination

Publication Type: Original research article

Preferred Citation: Ghoshal, Raj. 2019. “Flawed Measurement of Hiring Discrimination against African Americans.” *Sociation*, 18(2), 36-46.



This work is licensed under a [Creative Commons Attribution-Noncommercial 2.0 Generic License](https://creativecommons.org/licenses/by-nc/2.0/)

Introduction

African Americans experience higher unemployment and receive lower average wages than members of other racial groups in the United States. Many sociologists and economists (e.g., Pager 2003; Neumark 2012) have argued that hiring discrimination by employers plays a role in this gap. However, some economists (e.g., Becker 1957; Farkas and Vicknair 1996; Neal & Johnson 1996; Heckman 1998) counter that racial disparities in employment and income arise from on-average group differences in traits associated with race, such as educational attainment, rather than from employers’ reactions to applicants’ race. This

disagreement matters: efforts to reduce discrimination by employers are either necessary or a harmful diversion from addressing more fundamental causes of inequality like educational disparities (Heckman 1998), depending on who is correct.

In the past fifteen years, numerous experimental audit studies of the hiring process have sought to resolve this debate (Baert 2018). In a typical audit, researchers create fictitious identities with distinctive names that signal candidates’ racial backgrounds. They use these identities to respond to hundreds or thousands of job openings. Resume quality is experimentally held constant across different identities, so any significant variation in callback rates

can be attributed to discrimination. Recent audit studies in the United States consistently find that employers often still discriminate against African American job seekers (Quillian et al. 2017).

However, Darolia et al. (2016) argue that the claim of discrimination by employers is on shaky ground, as most audit studies of hiring and race share a methodological flaw. In their argument, most existing research incorrectly conflates signals of race with signals of class background, since distinctively black names are found most often among less-educated, lower-income families (Fryer and Levitt 2004). As an alternative, they use the first names “Ryan” and “Chloe” combined with two typically black last names, “Washington” and “Jackson,” to signal African American identity, and send 9,000 resumes to employers. They find equal callback rates across races. Recognizing that their “black names” might not be read as intended, Darolia et al. nonetheless argue that employers’ “error rate” in attributions of race would need to be “nearly 60 percent” (855), which they view as unlikely, for prior findings to stand. They suggest that the racial discrimination in hiring found by prior studies (e.g., Bertrand and Mullainathan 2004) may be either an artifact of race/class conflation in names or may have declined significantly over the past decade and a half (856).

Some recent work (e.g., Gaddis 2017) has cast doubt on Darolia et al.’s methodology, and most evidence suggests that discrimination in hiring persists. Nonetheless, further attention to their evidence and core claims is warranted, for two reasons. First, their work raises a serious methodological point: name selection in audit studies is a delicate balance, as clearly signaling race can be in tension with realistically representing a group’s name patterns. Studies that use overly stereotypical names might yield invalid findings, while using names whose race signals are weaker than intended also has pitfalls. A better understanding of how names used shape results attained can help us gauge audits’ validity.

Second, Darolia et al. (2016)’s study has clear implications for how government and employers should address racial inequality, and due to its policy relevance received significant media attention. Headlines to news coverage of the study included “New Research Refutes Famed Study of Hiring Bias,” (Elejalde-Ruiz 2016a), “Resumes with Black, White, and Hispanic Names Treated the Same” (Elejalde-Ruiz 2016b), “Race and Gender May Not Affect Employer Interest in Resumes” (Science Daily 2016), and “Black or Latino Surnames Don’t Actually Hurt Job Applicants” (Fenske 2016). In news reports, the authors were at times cautious about the implications

of their findings (Elejalde-Ruiz 2016b) but at other times appeared to suggest that racial discrimination in hiring may have been eliminated (Fenske 2016; University of Missouri-Columbia 2016). Methodological scrutiny of highly visible, policy-relevant scholarship is essential because such scholarship can impact real-world decision-making, as it did in America’s debate about marriage equality (Cohen 2012; Gates 2012). To better understand Darolia et al.’s policy implications, it is important to adjudicate between three competing possibilities: (1) that Darolia et al. were correct and discrimination in hiring is low or nonexistent; (2) that their methods were sound but their findings exceptional (that is, externally invalid); or (3) that their methods were so flawed that their study should be disregarded.

For these reasons, this study presents as direct a test as possible of Darolia et al. (2016). Building on Gaddis (2017), I conduct a survey of 1,050 American adults that asks respondents to indicate the most probable race of someone with a particular name. I test the four exact first/last name combinations used as “black names,” along with ten other name combinations. I find that most respondents did not read the four names as Darolia et al. intended. The overall error rate for those name combinations is 59.7 percent (or “nearly 60 percent”), far higher than for any other name combinations tested. The error rate is nearly unchanged (and in fact, slightly higher) when I restrict the sample to the types of individuals most likely to make hiring decisions. Additional analyses show that these findings are not an artifact of my sample, but due to problems with the name signals themselves. The level of error found invalidates Darolia et al.’s core claim. I show that prior conclusions of discrimination are not just unchallenged but bolstered by their study.

Literature Review

The U.S. civil rights movement led to advances in legal equality for African Americans, including landmark federal legislation around voting rights, civil rights, equal employment opportunity, and fair housing (Chafe 1981). However, de jure equality has not yet yielded *facto* equality. Over fifty years after the movement’s heyday, black Americans continued to be disadvantaged on important indicators, including those involving the labor market such as unemployment, poverty, wealth, and income (Santoro 2015). While there are many causes of ongoing racial inequality, employment is one of the most crucial institutions to consider. Employment shapes many other aspects of life such as the ability to buy a house, pay for medical emergencies, and send one’s children to college. Understanding and remedying African Americans’ ongoing disadvantage in the labor market

is a necessary and important step in addressing racial inequality.

Although scholars agree that black Americans fare worse than most other groups in the labor market, they debate the roots of this pattern. In particular, one dispute concerns whether employers engage in racial discrimination against black job candidates when hiring (Neumark 2018). Classical economic theory expects racial discrimination by employers to be rare due to its economic irrationality: employers who discriminate on grounds unrelated to job performance are depriving themselves of valuable workers, reducing their own competitiveness (Heckman 1998; Becker 2010). However, some scholars have countered that regardless of what classical theory expects, there is ample empirical evidence of hiring discrimination (Baert et al. 2015; Pager 2016; Neumark and Rich 2019; Quillian et al. 2019). They also point out that employers may indulge in “taste-based discrimination” or unconscious discrimination even if it hurts their profits, and that statistical discrimination in hiring might be profit-neutral or profitable in some cases (Guryan and Charles 2013; Neumark 2018). In sum, classical economic theory sees black Americans’ lower economic attainment as fully or near-fully due to lower education and experience (Farkas and Vicknair 1996; Neal and Johnson 1996; Fryer and Levitt 2004; Becker 2010 [1957]), while some economists and sociologists see racial discrimination by employers, whatever its motivations, as an important additional force shaping hiring decisions.

Race and educational attainment are correlated, so for decades, this debate proved challenging to resolve. However, one powerful tool was developed to control for the race-human capital correlation and assess discrimination directly: audit studies. In audit studies (Pager 2007; Gaddis 2018), civil rights agencies and researchers could hire individuals of different races, arm them with equal-quality resumes, train them to present equally well in person, and send them to apply for the same set of jobs in person or over the phone. Scholars would then monitor callback rates or expressions of interest by employers using different voicemail boxes or P.O. boxes for each candidate. Nevertheless, researching discrimination using this approach was expensive, time-consuming, and potentially error-prone, since knowing whether different candidates came across equally well was difficult (Heckman 1998). While some audit studies suggested hiring discrimination (Pager 2003), these studies were relatively rare and had to be interpreted cautiously.

Since the mid-2000s, the online shift of job search processes has greatly facilitated scholars’ ability to research hiring discrimination. Now, researchers can

choose names characteristic of people of different races, create email accounts associated with these names, and use these email accounts to send out hundreds or thousands of equal-quality resumes to a broad set of employers (Crabtree 2018; Gaddis 2018). Because resume quality does not vary across the different names tested, differences in response rates can be attributed to employers’ reactions based on perceived race. In the last 15 years, hundreds of studies have employed this method of studying racial discrimination in hiring across dozens of countries (Baert 2018; Quillian et al. 2019). The most evident finding about the United States in this line of work is that equal-quality resumes fare worse when associated with a typically African American name than with a typically white name. Many scholars have concluded that employers’ discriminatory decisions are one important influence on ongoing racial inequality.

Darolia et al. (2016) challenged this viewpoint. In their argument, most audit studies share an important methodological flaw. If the names used to signal a candidate as African American are disproportionately found among people of lower socioeconomic class origins, what appears to be discrimination based on race may, in fact, be motivated by cues around class background. The problem is magnified if the names used to signal black identity are unusually stereotypical or signal lower-SES origins more than most black people’s names do. For instance, “Lakisha” is typically an African American name, but it is not a typical African American name (Fryer and Levitt 2004). In Darolia et al.’s argument, perhaps candidates with less stereotypical names encounter fairer treatment than candidates with the kinds of names most often used in audit studies.

Darolia et al. therefore conducted their own audit study of hiring for 3,000 job openings across the United States. They broke with prior studies’ approach to choosing names by using the first names Ryan and Chloe in combination with the last names Washington and Jackson for their African American candidates. While they described the last names they use as “strong indicators of race” (90 percent of Washingtons and 53 percent of Jacksons are African American; U.S. Census 2016), they characterized the names “Ryan” and “Chloe” simply as “male-sounding” and “female-sounding.”

After sending out 9,000 resumes, they found only minimal differences in callback rates by apparent race and attributed their finding of no discrimination to having avoided mixing race and class signals. They acknowledged that some employers might not recognize “Ryan Washington,” “Chloe Washington,” “Ryan Jackson, and “Chloe Jackson” as names intended to signal black candidates. However, they argued that the “error rate” in perceiving these names

as African American would need to be “nearly 60 percent” for Bertrand and Mullainathan’s (2004) finding that black candidates needed to send out 1.5 times as many resumes as white candidates to receive proportionate response to be correct. They also suggested that racial discrimination in hiring may have declined in recent years.

Darolia et al. (2016) ran counter to a great deal of prior work. Indirect challenges followed from several sources. Concurrent and successive audit studies, including one that equalized educational credentials between white and black job applicants, found discrimination against black job applicants (Gaddis 2015). A meta-analysis found no evidence of a downward trend in discrimination (Quillian et al. 2017). Videos recorded on cell phones and circulated on social media provided ample anecdotal evidence of ongoing adverse treatment of black Americans in numerous situations (Wootson 2018). Most directly, Gaddis (2017) used a convenience sample to test how Amazon Mechanical Turk users (Chandler and Shapiro 2016) perceived the race of various names. He found that “Ryan” and “Chloe” were both perceived most often as white names. This finding implies the possibility that the signals intended from Darolia et al.’s four “black names” might have been misread quite often, perhaps even “nearly 60 percent” of the time. If true, this would invalidate their study. This study investigates this possibility in greater depth.

Methods and Data

To investigate how people read Darolia et al.’s names, I designed an online survey module and fielded it in the fall of 2017. The module was distributed to a national quasi-probability sample as one component of a more extensive survey that addressed many different topics; respondents had previously opted into receiving online surveys in exchange for compensation. In total, 1,050 respondents participated. Respondents answered various demographic questions before turning to modules, including my module about race and name perceptions. Demographic items included respondents’ race (white, black, Hispanic, Asian, or other), age, gender, Census region, political party (Democrat, Republican, independent, or other), employment status (full-time, part-time, or not currently employed), education level (in six categories), and income (in \$25,000 increments, up to “\$100,000 a year or more”). I conducted analyses on the full sample and also on a subsample designed to more closely approximate traits of people making hiring decisions, described below.

Overall, respondents reflected key US adult population demographics. Fifty-one percent of respondents were female. The mean age was 44, and

respondents ranged from 18 to 81 years. Respondents’ location by Census region matched overall US patterns, with 37 percent hailing from the South, 23 percent from the West, 21 percent from the Midwest, and 18 percent from the Northeast. Sixty-eight percent were white, 13 percent were Hispanic, 12 percent were African American, 3 percent were Asian, and 3 percent identified as other-race. This racial distribution approximates the US adult population at the time Darolia et al.’s study was conducted (66 percent non-Hispanic white, 15 percent Hispanic, 12 percent African American, 5 percent Asian; 2 percent other in 2013-14) and likely differs even less from the adult English-speaking population at that time. Though people who complete surveys online may vary from other people in some ways, on balance, these differences are slight and unlikely to skew inference around the patterns I identify, a point I address further in the discussion. Only respondents who opted in were counted, so I do not report a response rate.

In the module, all survey respondents were given a list of 14 names and asked to categorize each name as most likely belonging to someone who is White; Black or African American; Hispanic or Latino; or Asian, Indian, or Middle Eastern. Specifically, respondents were told, “First, for each of the 14 names given below, please choose the racial/ethnic group you’d guess someone with that name is most likely a member of. Please do not look up information on any name; we’re interested solely in your first impressions. Names are randomized, so don’t worry if you end up with some empty or some very full categories.” Every respondent saw the same 14 names. Respondents could choose only one of the four options per name and had to choose an option to continue. Names included Darolia et al.’s four “African American names” along with two names they had used to signal white identity, two of their “Hispanic names,” and six other names that did not appear in their study but which I included for comparison. In the second part of the study, respondents classified additional names following a slightly different design.

After all respondents completed the survey, I examined their classifications of each name. I calculated the error rate for classification of each name compared to the signal intended, and for Darolia et al.’s names as a group. Because people making hiring decisions may differ from other adult Americans on variables such as employment status and income, I additionally conducted an analysis among a subsample with greater demographic similarity to most hiring decision-makers. For this subsample analysis, I selected only respondents working full time, ages 30-65, with some college or more, and with annual incomes of \$50,000 or higher. Given these variables’ correlations with race and gender, the subsample of

277 respondents was whiter (75 percent white) and more male (61 percent male) than the full sample, realistically reflecting how race and gender intersect with work authority (Quadlin 2016). Methodology in the subsample analysis mirrored that in the larger analysis.

I also tested the explanatory power of respondents' race, gender, age, region, employment status, education, income category, and political party on name identification patterns among the full sample. For analyses of gender, region, education, employment, and income category, I retained the original survey categories. I consolidated respondents who were neither white nor black into a third race category and grouped age into six clusters, with those 30 or younger as the lowest group and those 70 or higher as the highest group. I consolidated political independents and other-party identifiers into a single group.

Results

Table 1 presents respondents' attributions of race to Darolia et al.'s four names of African American applicants. As seen in the top half, among the full sample, the names Ryan Jackson and Ryan Washington were identified as African American less than a third of the time (32.1 percent and 31.2 percent, respectively), and thought to belong to white individuals nearly two-thirds of the time (63.2 percent and 64.5 percent of the time). Chloe Jackson and Chloe Washington were read as intended somewhat more often; respondents thought of these names as belonging to black individuals about half the time. However, the four "black names" were perceived as black only 40.3 percent of the time overall, yielding an "error rate" of 59.7 percent among the full sample.

Do individuals similar to those who have more authority over hiring perceive names and race differently than others? I present results from the subsample analysis in the bottom half of Table 1. While differences are small, those in the subsample were very slightly less likely to read the four names as intended than the overall sample: they read the four names as black only 39.6 percent of the time, yielding an error rate of 60.4 percent. Experimenting with variations in subsample structure (for instance, changing the age range selected, or limiting the subsample to those with a 2-year degree or higher) did not change the overall pattern; the error rate for these names remained about 60 percent.

The evidence presented so far is sufficient to show that about 60 percent of Darolia et al.'s subjects may well have misread their intended racial signal. It is also instructive to compare error rates for the four African

American names they used to other names and to consider possible effects of respondents' race. While Gaddis (2017a; 2017b) shows that people usually infer race as intended in audit studies, here I present directly comparable data. I compare error rates for Darolia et al.'s four "African American names" to the other ten names I asked all survey respondents to consider. These ten names included two of the four white names (Brian Thompson and Megan Anderson) and two of the four Hispanic names (Isabella Hernandez and Carlos Garcia) used in their study, along with three names I expected would be read as typically Asian (Jia Chang, Riya Patel, and Rahul Singh) and three names I expected would be read as typically African American (Deandre Jefferson, Tyra Booker, and Shanice Banks).

Figure 1 presents the results of this analysis. The overall error rate for these other ten names as a group was only 19.2 percent, significantly (and substantially) different than the nearly 60 percent error rate for Darolia et al.'s names ($p < .0001$). This 48-point gap far exceeds the gaps in identification rates between any other clusters, as average error rates in the non-Darolia et al. clusters were 16 percent for the two white names, 19 percent for the five Asian and Hispanic names, and 22 percent for the other three black names. On the respondent side, whites' error rate of 60.0 percent on the Darolia et al. names was nearly identical to that of Asians, Hispanics, and those of other races at 60.9 percent; black respondents were slightly but not significantly ($p = .09$) more accurate than other groups, with an error rate of 56.1 percent. Whites, blacks, and people of other races all made 2.5 to 3.5 times as many errors in identifying Darolia et al.'s names, compared to the other names. These patterns suggest the high error rate for Darolia et al.'s names was not due to flaws particular to my sample (since sample members identified other names with reasonably high accuracy), nor did it cut equally against all the names they used. Rather, the flaws were solely in the names they intended to be read as African American.

To gain additional insight into whether Darolia et al.'s subjects differed from my sample in any critical ways and into forces shaping perceptions, I tested whether name identification varied by respondents' race, gender, age, region, educational attainment, employment, income, and political party. In 30 out of 31 demographic subgroups tested, fewer than half of respondents identified the names as intended. The only group to read Darolia et al.'s names as black at over 50 percent was respondents age 70 and over. Bivariate analyses found that older respondents, men, and political independents/others had higher "correct" identification rates than younger people, women, Democrats, and Republicans. However, the effect sizes were tiny for everything but age, where each

additional decade yielded about 5 percent greater accuracy ($p < .001$). All other variables were insignificant.ⁱ In a multivariate analysis, only age retained its significance. While the age effect is interesting, the results still show that those making hiring decisions would not usually identify Darolia et al.'s names as intended. Indeed, the overall pattern here of a single notable effect among all the variables tested is consistent with what might occur by chance (Simmons, Nelson, and Simonsohn 2011).

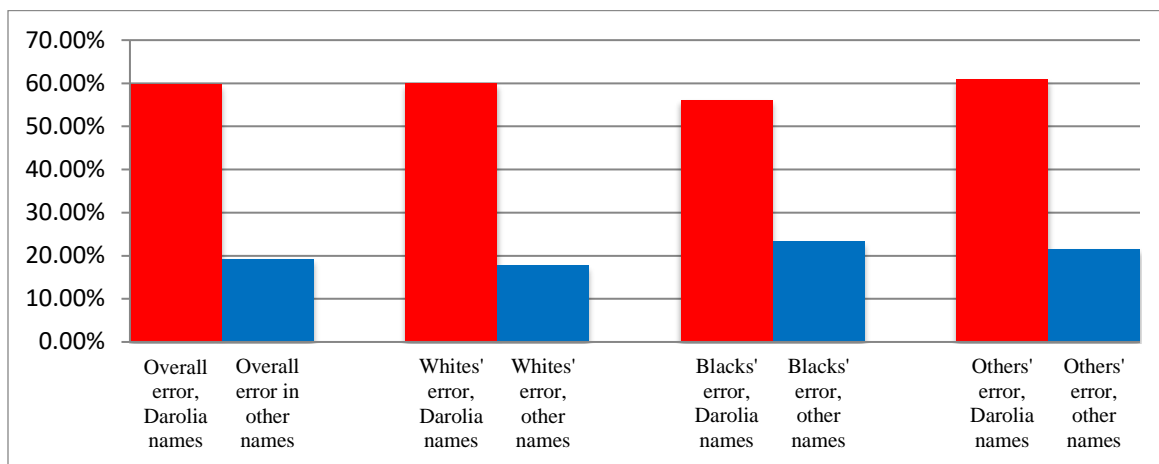
I raise one final point on why respondents may have perceived names as they did. About 53 percent of Jacksons and 90 percent of Washingtons, but only about six percent of Ryans and eight percent of Chloes, are African American (U.S. Census 2016; Gaddis 2017; Gaddis and Ghoshal 2019). Therefore, the two last names Darolia et al. (2016) used for African American candidates would, on average, belong to a

black person about 71.5 percent of the time, while their two first names would do so about seven percent of the time. Respondents who unconsciously perceived these percentages accurately and gave equal weight to the first and last name in determining race would, therefore, perceive Darolia et al.'s four names as belonging to black individuals $((71.5 + 7) \times 0.5)$ percent, or 39.25 percent, of the time. This yields an "error rate" of 60.75 percent. To be clear, my claim here is not that most people engage in this process of calculation, nor that this method of calculation is correct.ⁱⁱ It is merely that nothing is shocking in the high error rate. Instead, it is dismissing the possibility of a very high error rate out of hand, as Darolia et al. appear to do, that is unreasonable.

Table 1. Perceived Race of Darolia et al.'s "African American Names."

	Black	White	Hispanic	Other
<i>By full sample:</i>				
Ryan Jackson	32.1% (337)	63.2% (664)	3.4% (36)	1.2% (13)
Chloe Jackson	50.2% (527)	43.2% (454)	5.0% (52)	1.6% (17)
Ryan Washington	31.2% (328)	64.5% (677)	3.0% (31)	1.3% (14)
Chloe Washington	47.8% (502)	48.7% (511)	2.1% (22)	1.4% (15)
Average	40.3%	54.9%	3.4%	1.4%
<i>By subsample:</i>				
Ryan Jackson	27.1% (75)	67.5% (187)	4.3% (12)	1.1% (3)
Chloe Jackson	48.4% (134)	44.8% (124)	5.8% (16)	1.1% (3)
Ryan Washington	33.9% (94)	62.5% (173)	2.9% (8)	0.7% (2)
Chloe Washington	49.1% (136)	48.0% (133)	2.6% (7)	0.4% (1)
Average	39.6%	55.7%	3.9%	0.8%

Figure 1: Error Rate in Name Attributions for Darolia et al. Names and Ten Other Names.



Discussion and Conclusion

Numerous prior audit studies have found that employers engage in hiring discrimination. However, Darolia et al. (2016; Fenske 2016) argued that employers' equal treatment of typically-white names and the "black names" Ryan Washington, Ryan Jackson, Chloe Washington, and Chloe Jackson might indicate the waning of such discrimination. This study directly tested a crucial part of Darolia et al.'s methodology and found it wanting. Specifically, the survey data presented here reveal an alternative explanation for the equal treatment Darolia et al.'s names received: neither adult Americans generally, nor Americans demographically more likely to make hiring decisions, perceive the names as belonging to black individuals. Their study, therefore, provides no valid evidence that racial discrimination in hiring has been eliminated. In the remaining paragraphs, I consider two limitations and the broader implications of the present study.

One limitation is that I did not measure name perceptions directly among people making hiring decisions. Perhaps those who review resumes are more adept than other people at associating names with the race they typically signify. If Darolia et al.'s targets read the four "black names" names as intended that would indeed suggest a decline in discrimination. However, very-white first names and very-black last name combinations are scarce. There is, therefore, no reason to expect that recruiters and hiring managers

would have clear impressions of race from these name combinations. Indeed, this study's subsample of respondents most likely to be involved in hiring was slightly less likely to read the names as intended than the full sample. This may be partly because African Americans are usually better able to identify typically-black names than others (Gaddis 2017) but are underrepresented among better-paid and higher-authority workers (Quadlin 2018). Additionally, given that even black respondents' "correct" identification of these four names, in particular, was extremely low, it is also likely that the four names used were unrealistic and sent very weak signals.

A second limitation is that this study drew on an online sample rather than more traditional survey methods. This could skew inference if respondents simply "mashed buttons" rather than giving real responses, or if respondents are unusual in any important ways relevant to the topic. However, respondents correctly identified all other names at very high rates, showing they took the survey seriously. The sample was designed to match the American adult population on key demographic variables like race,

gender, and region, bolstering representativeness. While Hispanics and Asians were slightly underrepresented and whites slightly overrepresented in the overall sample, this had no effect on the results, as Hispanics and Asians' error rate was less than one percent different than whites (and in fact was very slightly higher). Even if people who opt into internet surveys may differ in some ways from American adults generally, there is no reason this difference would skew identification rates for four names in particular but not ten others. Finally of note here, given the difficulty of building a representative sample of people making hiring decisions across a wide array of industries and locations, prior samples used to test name perceptions have generally been very casually drawn convenience samples or unrepresentative samples of Amazon Mechanical Turk users (Chandler and Shapiro 2016; Gaddis 2017). The samples here are a step forward.

The main implication is that Darolia et al. fail to show that racial discrimination in hiring is no longer a significant concern. Perhaps more strikingly, this study shows that their data, when correctly interpreted, actually strengthen prior findings of discrimination. As they state, an error rate in name attribution of "nearly 60 percent" reveals a level of discrimination "similar to what is found by Bertrand and Mullainathan" (Darolia et al. 2016: 855). Bertrand and Mullainathan found those job candidates with typically-black names needed to send out about 1.5 times as many resumes as white-named candidates to receive a proportionate response. Therefore, a correct interpretation of Darolia et al.'s data does not merely discard their study; instead, it supports the opposite conclusion than the one they drew. Racial discrimination in hiring persists

Despite its flaws, Darolia et al. raise one point future research could productively address: how strongly names used in audit studies should signal race. After all, the five most common last names among African Americans are Williams, Johnson, Smith, Jones, and Brown. However, there are more white Americans than black Americans with each of these last names (author's calculation based on Census data), so they are not distinctive. Similarly, distinctively "black-sounding" first names are relatively rare even among black Americans. Will black Americans with "white-sounding" or non-stereotypical names face the same treatment as those with racially distinctive names if they make it to the in-person stage? If we assume that both the motivations for and the extent of discrimination at the in-person stage of hiring mirrors that at the resume-filtering/callback stage, then the answer is yes, and resume studies using only very distinctively black

names provide useful information on how all black Americans will be treated.

Alternatively, hiring processes might differ across stages. For instance, resumes may be screened by one person acting alone while in-person interviews are conducted by a group who discuss their deliberations, or vice versa. The primacy of conscious versus unconscious motivations for discrimination, the extent to which decisions need to be justified, and other factors may vary across stages, yielding less clarity on how African Americans with non-distinctive names will fare across multiple stages. Some comparative research testing the impacts of race at different stages of a single study has begun, and early evidence points to discrimination at all stages of the hiring process (Quillian, Lee, and Oliver 2018). Related work suggests that anti-black name-based discrimination is not reducible to class cues (Gaddis 2017) and that decision-makers may focus on names' race signals much more than on their class signals (Butler and Homola 2017). These points are suggestive that racial exclusion of black candidates with non-typically black names will occur at later stages of hiring and that audit studies' use of unusually distinctive names likely does not invalidate their results, but more work is needed to understand how racial signals operate at different stages of the hiring process. Expanding this line of work offers the chance to better understand the motivations, processes, and scope of unequal treatment directed at black Americans, a necessary step toward diminishing persistent discrimination and inequality.

References

- Baert, Stijn. 2018. "Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005." In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, edited by S. Michael Gaddis, 63-77. Cham, Switzerland: Springer.
- Baert, Stijn, Bart Cockx, Niels Gheyle, and Cora Vandamme. 2015. "Is There Less Discrimination in Occupations Where Recruitment Is Difficult?" *ILR Review* 68(3): 467-500. doi: [10.1177/0019793915570873](https://doi.org/10.1177/0019793915570873).
- Becker, Gary S. 2010 [1957]. *The Economics of Discrimination*. Chicago: University of Chicago.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4): 991-1013.
- Butler, Daniel, and Jonathan Homola. 2017. "An Empirical Justification for the Use of Racially Distinctive Names to Signal Race in Experiments." *Political Analysis* 25(1):122-30.
- Chafe, William H. 1981. *Civilities and Civil rights: Greensboro, North Carolina, and the Black Struggle for Freedom*. New York: Oxford University Press.
- Chandler, Jesse, and Danielle Shapiro. 2016. "Conducting Clinical Research Using Crowdsourced Convenience Samples." *Annual Review of Clinical Psychology* 12:53-81.
- Cheng, Simon, and Brian Powell. 2015. "Measurement, Methods, and Divergent Patterns: Reassessing the Effects of Same-sex Parents." *Social Science Research* 52: 615-26.
- Cohen, Philip. 2012. "Regnerus Study Controversy Guide." Blog post at *Family Inequality*, August 15. Last retrieved December 18, 2019 (<https://familyinequality.wordpress.com/2012/08/15/regnerus-study-controversy-guide/>).
- Crabtree, Charles. 2018. "An Introduction to Conducting Email Audit Studies." In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, edited by S. Michael Gaddis, 103-117. Cham, Switzerland: Springer.
- Darolia, Rajeev, Cory Koedel, Paco Martorell, Katie Wilson, and Francisco Perez-Arce. 2016. "Race and Gender Effects on Employer Interest in Job Applicants: New Evidence from a Resume Field Experiment." *Applied Economics Letters* 23(12): 853-56.
- Elejalde-Ruiz, Alexia. 2016a. "New Research Refutes Famed Study of Hiring Bias Based on Names." *Minneapolis Star-Tribune*, May 14. Last retrieved December 19, 2019 (<http://www.startribune.com/new-research-refutes-famed-study-of-hiring-bias-based-on-names/379747291/>).
- Elejalde-Ruiz, Alexia. 2016b. "Hiring Bias Study: Resumes with Black, White, Hispanic Names Treated the Same." *Chicago Tribune*, May 4. Last retrieved December 19, 2019 (<https://www.chicagotribune.com/business/ct-bias-hiring-0504-biz-20160503-story.html>).
- Farkas, George, and Keven Vicknair. "Appropriate Tests of Racial Wage Discrimination Require Controls for Cognitive Skill: Comment on Cancio, Evans, and Maume." *American Sociological Review* 61: 557-60.
- Fenske, Sarah. 2016. "Black or Latino Surnames Don't Actually Hurt Job Applicants, Mizzou Study Finds." *Riverfront Times*, April 27. Last retrieved December 19, 2019 (<https://www.riverfronttimes.com/newsblog/2016/04/27/black-or-latino-surnames-dont-actually-hurt-job-applicants-mizzou-study-finds>).

- Fryer, Roland G. Jr., 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. 2015. "Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market." *Social Forces* 93(4): 1451-79.
- Gaddis, S. Michael. 2017. "How Black are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." *Sociological Science* 4: 469-89.
- Gaddis, S. Michael. 2018. *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. Cham, Switzerland: Springer.
- Gaddis, S. Michael, and Raj Ghoshal. 2019. "Dynamic Racial Triangulation: Examining the Racial Order using Two Experiments on Discrimination among Millennials." SSRN. Last retrieved December 19, 2019 (<http://dx.doi.org/10.2139/ssrn.3022208>).
- Gates, Gary. 2012. "Letter to the Editors and Assistant Editors of Social Science Research." Letter posted online at *Family Inequality* blog. Last retrieved December 19, 2019 (<https://familyinequality.files.wordpress.com/2012/08/ssr-gates-letter.pdf>).
- Guryan, Jonathan, and Kerwin Kofi Charles. 2013. "Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots." *The Economic Journal* 123(572): F417-F432.
- Heckman, James. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12: 101-16. doi:[10.1257/jep.12.2.101](https://doi.org/10.1257/jep.12.2.101).
- Neal, Derek, and William Johnson. 1996. "The Role of Pre-market Factors in Black-White Wage Differences." *Journal of Political Economy* 104(5):869-95.
- Neumark, David. 2012. "Detecting Discrimination in Audit and Correspondence Studies." *Journal of Human Resources* 47(4): 1128-57.
- Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56(3): 799-866.
- Neumark, David, and Rich, Judith. 2019. "Do Field Experiments on Labor and Housing Markets Overstate Discrimination? A Re-examination of the Evidence." *ILR Review* 72(1): 223-52. <https://doi.org/10.1177/0019793918759665>
- Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108(5): 937-75.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *The ANNALS of the American Academy of Political and Social Science* 609(1): 104-33. doi:[10.1177/0002716206294796](https://doi.org/10.1177/0002716206294796).
- Pager, Devah. 2016. "Are Firms that Discriminate More Likely to Go Out of Business?" *Sociological Science* 3: 849-59.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34: 181-209.
- Quadlin, Natasha. 2018. "The Mark of a Woman's Record: Gender and Academic Performance in Hiring." *American Sociological Review* 83(2): 331-60. Doi:[10.1177/0003122418762291](https://doi.org/10.1177/0003122418762291).
- Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. 2017. "Meta-Analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring over Time." *Proceedings of the National Academy of Sciences* 114 (41): 10870-75. doi:[10.1073/pnas.1706255114](https://doi.org/10.1073/pnas.1706255114).
- Quillian, Lincoln, Anthony Heath, Devah Pager, Arnfinn H. Midtbøen, Fenella Fleischmann, and Ole Hexel. 2019. "Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring." *Sociological Science* 6: 467-96.
- Quillian, Lincoln, John Lee, and Mariana Oliver. 2018. "Who Gets the Callback? Who Gets the Job? A Comparison of Callback and Job Offer Outcomes in Field Experiments of Hiring." Presentation at the 2018 Meeting of the Southern Sociological Society, New Orleans, LA.
- Santoro, Wayne A. 2015. "Was the Civil Rights Movement Successful? Tracking and Understanding Black Views." *Sociological Forum* 30: 627-47.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22(11): 1359-66.
- U.S. Census Bureau. 2016. "Frequently Occurring Surnames from the 2010 Census." Last retrieved December 19, 2019 (https://www.census.gov/topics/population/genalogy/data/2010_surnames.html).
- University of Missouri-Columbia. 2016. "Race and Gender May Not Affect Employer Interest in Resumes." *ScienceDaily*, April 26. Last retrieved December 19, 2019 (<https://www.sciencedaily.com/releases/2016/04/160426162606.htm>).
- Wootson, Cleve Jr. 2018. "A Black Yale Student Fell Asleep in Her Dorm's Common Room. A White

Student Called Police.” *The Washington Post*,
May 11. Last retrieved December 19, 2019
(<https://www.washingtonpost.com/news/grade->

[point/wp/2018/05/10/a-black-yale-student-fell-asleep-in-her-dorms-common-room-a-white-student-called-police/?utm_term=.fb32cbf0fc01](https://www.washingtonpost.com/news/grade-point/wp/2018/05/10/a-black-yale-student-fell-asleep-in-her-dorms-common-room-a-white-student-called-police/?utm_term=.fb32cbf0fc01)).

ⁱ Alternate specifications, such as maintaining age as a fully continuous variable, did not meaningfully change the results.

ⁱⁱ More Americans have the last names Washington or Jackson than have the first names Ryan or Chloe, so simple averaging is not the best method to calculate combined name prevalence. Taking the names’ different frequencies into account, and if one assumes that black and non-black Americans with the last names Washington and Jackson choose the first names Ryan and Chloe at rates similar to other people of their own race, then Darolia *et al.*’s average “black name” would belong to a black individual about 58 percent of the time. Of course, very-white first names and very-black last names rarely overlap, and it is unclear whether naming patterns among non-blacks with typically black last names are different than other non-blacks, so the actual prevalence of these names may be different.