Best Man For The Job? How Bias Affects Hiring

By Isaac Butler and Rachel Godsil

We all like to believe that our qualifications determine whether we are invited for job interviews and that our performance dictates whether our employers think well of us. Obviously, employers are better off if they hire the most qualified applicants. And of course, the goals of business and government goals are furthered when managers rate employees according to how well they perform their jobs. Recent research strongly suggests, however, that implicit bias affects both whether people are invited to interview for jobs, regardless of their qualifications, as well as how their performance is assessed.

1. Hiring – Who Gets the Interview?

A 2004 study by Marianne Bertrand and Sendhil Mullainathan demonstrates that employers’ implicit biases play a significant role in the interview process. By sending out resumes with identical qualifications, half of which had names commonly associated with African Americans – “Lakisha” and “Jamal” — and half of which had names associated with whites – “Emily” and “Greg” in response to job interviews in Chicago and Boston, they discovered a 50% gap in callback rates between white- and black- named resumes. Whites were invited for one interview for every ten resumes they sent, while blacks had to send fifteen for a single interview. Even more troubling, they found that “racial gaps in callback[s] are statistically indistinguishable across all categories: Federal contractors, who are thought to be more severely constrained by affirmative action laws, do not treat African-American resumes more preferentially; neither do large employers or employers who explicitly state that they are `Equal Opportunity Employers.'”

Bertrand and Mullainathan went to great lengths to control for factors other than race. First, they based their “dummy” resumes on existing resumes they found on job search websites. Next, they created two different pools of resumes. Half of them were of higher quality (possessing more work experience, summer employment while in college, extra computer skills, certification degrees, foreign language skills, honors or some military service) and half were of lower quality. Controlling for geography, they performed the test in two different cities (Boston and Chicago) and listed addresses from both affluent and poor neighborhoods. For each job listing they found, they sent out four resumes, two of higher quality and two of lower quality. They randomly assigned one “black name” and one “white name” to each group, so each prospective employer received one black high quality, one white high quality, one black lower quality and one white lower quality resume. Accordingly, a resume was as likely to end up with a white or black name, regardless of address, education level or any other factor. They also set up dummy voice mail and e-mail accounts to receive the callbacks.

As a result of these different variables, Bertrand and Mullainathan were able to extract a variety of conclusions from their study. Fundamentally, whites were called back on average 9.65% of the time, while African Americans were called back at a rate of 6.45%. While “equal treatment” (both races getting the same result) did occur roughly 88% of the time — in the sense that 88% of all resumes, black and white, were rejected — it is the resumes that are accepted that count for true equal treatment. “Higher-quality resumes receive[d] more callbacks” but resumes with black-sounding names did not experience the same degree of quality bump that white-sounding resumes did. A high quality white resume was called back 11% of the time versus 8.5% for a low quality white resume. For African Americans, however, the rates were 6.7% for a high quality resume and 6.2% for a low quality one. In other words, a low quality white applicant was more

likely to be invited to interview than a high quality black applicant. And for African Americans, there was only a half a percent difference in the likelihood of being invited to interview based upon education level, while for whites, the difference was 3.5%. The question simply must be asked: what incentives for educational achievement does this treatment invite?

Bertrand and Mullainathan were particularly conscious of the potential for race and class to be conflated. Accordingly, they designed their study to determine whether race is actually a signifier for class and that class issues, not race alone, are what hold African Americans (and poor whites) back. They analyzed this hypothesis through two different lenses. First, they cross checked the resumes with the fake applicant’s address. Some addresses (again, randomly assigned) were from “good” neighborhoods, and some from “bad” areas of town. They found that “there is no evidence that African-Americans benefit any more than whites from living in whiter, more educated Zip Code[s].” They also double-checked the African American and White names for class identification. “We were able to obtain birth certificate data on mother’s education… for babies born in Massachusetts between 1970 and 1986.” Using education-level as a proxy for class, they determined the average class background of the various names. While white names do, on average, indicate a higher class background level than black names, within each racial group, the class indication of the name made little difference. “Leroy” (a black-associated name) got called back 9.4% of the time, but his name indicates a lower class-background-by-proxy than “Rasheed” who got called back 3% of the time. The obvious difference between Leroy and Rasheed as names is their otherness, and these findings suggest that it is the otherness of the name that is prejudicial.

The fact that hiring discrimination persists is one reason many would argue that affirmative action programs are still necessary – they serve to create an incentive for employers to overcome their implicit bias and to take a second look at candidates of color. However, the authors pointed out that many jobs are filled through social networks rather than through blind resume submission. If blacks are continuing to face hiring discrimination and to have trouble simply getting to the interview phase of the hiring process, diversifying those social networks will continue to be a persistent, difficult problem to resolve. Despite the popular specter of “reverse discrimination”, this study shows exactly how prevalent racial bias in hiring actually is.

A very similar test in Sweden (Implicit Discrimination in Hiring: Real World Evidence by Dan-Olof Rooth) utilized an identical resume experiment with implicit and explicit bias tests of participating job recruiters. The bias most prevalent in Sweden is against Muslims, so the researchers used names associated with Arabs or Muslims instead of African Americans. They found that resumes with Arab/Muslim names were 10% less likely to be called in for an interview and that IAT scores indicating bias against Arabs directly correlated with the likelihood of a callback.

Rooth then followed up with many of the employers who had unwittingly taken part in the first half of the study. The employers filled out three different explicit measures of bias against Arabs and Muslims and then took an IAT that paired Swedish and Arab names with work- associated words such as “lazy”, “slow”, “efficient” and “hard-working”. Not surprisingly, “the IAT scores of the 193 recruiters participating in this study show that a very clear majority associate words signaling negative productivity… with belonging to the Arab/Muslim minority”.

When measuring the IAT scores and explicit measures against the actual outcomes of their previous resume submissions, they found clear evidence of the interplay between implicit bias and hiring of minorities. The slower the response time for pairing Arab/Muslim names and productive words (i.e. the higher the IAT score) the less likely the recruiter was to have called back an Arab/Muslim candidate. Furthermore, the IAT was a better measure of bias than the explicit measures. Only recruiters who would openly admit to preferring Swedish workers had explicit measure scores that correlated. Running the various numbers against each other, Rooth
concluded that fully half of the 10% gap between the callback rates of Swedish and Arab/Muslim dummy resumes was explainable by the existence of implicit bias.

B. Race, Gender and Performance Evaluation – Is Performance Being Evaluated?

Employers may either consciously or subconsciously justify their racial biases in hiring on the grounds that their most valued employees tend to be the same race. If, in their own judgment, the people who perform best are likely to be white, why shouldn’t they invite more white candidates to interview? It turns out, though, that there is good reason to suspect our evaluations of excellence. Instead of the actual performance of white employees exceeding the actual performance of minority employees, managers may very well instead be evaluating their employees based upon their stereotypes about who is most likely to be an effective worker.

In a recent study, AAV Researcher Advisors Jerry Kang and Nilanjana Dasgupta examine the interplay of stereotypes and hiring preferences, asking the question Are Ideal Litigators White? Their study found that stereotyping can lead people to rate the effectiveness of identical lawyers differently based solely on race.

Kang and colleagues married two different concepts to form the bedrock of their study. The first, which should be well-familiar by now, is that of implicit bias. The second is called Role Congruity Theory (RCT). Simply put, RCT says that people tend to use stereotypes when imagining various roles. The authors used the following example to explain how Role Congruity Theory works, and how it could impact implicit social cognition: A man is driving his son in a car when they get into a terrible car accident. The man is killed, the son is severely injured and is taken to the emergency room. The surgeon walks in and says, “I cannot operate on this boy, he is my son”. Who is the surgeon? Most people answering this question will go through many male guesses (step-dad etc.) before they get to the correct answer: The surgeon is the boy’s mother. This is because we tend to presume that the ideal “model” of a surgeon is a man.

Noting that job attributes are “not only strongly gendered, which has been the subject of previous studies... but also strongly racialized” the researchers sought to measure the effect of implicit bias on people’s evaluations of two hypothetical litigator candidates, one white and one Asian American. An Asian American was chosen because of stereotypes surrounding Asian Americans as “being deficient in interpersonal and social skills” as well as being “quiet and deferential”.

Sixty-eight subjects took a new IAT designed to measure White and Asian American associations with words that call to mind the ideal litigator or scientist. A second IAT was then administered focusing simply on implicit racial attitudes “or the degree to which [subjects] favored one racial group over another overall”.

Next, the subjects were shown a written transcript and an audio recording of two fictional depositions created by an experienced litigator. In each case, the litigator was deposing the opponent party. “The two depositions were created to be comparable in complexity, length... quality of the litigator’s performance, and ability to capture the listener’s interest”. The participants in the study initially saw a name and photograph for the litigator arguing the case. As you may have guessed, some participants were given a white litigator with a prototypically European sounding name (William Cole), while some were given an Asian American litigator with a prototypically Asian name (Sung Chang). After listening to the depositions, participants were asked to fill out an evaluation of the litigator including their perceived competence, perceived warmth, and willingness to hire or recommend the litigator to a family member or friend in need. Participants were then subject to an explicit self-report test in which they answered questions like “How ELOQUENT do you think WHITE AMERICAN litigators are?” on a scale of 1-7. They were also asked questions about their knowledge of various cultural stereotypes about whites and Asian American litigators.
As hypothesized, our vision of the ideal litigator is indeed racialized. Results on the litigator/scientist IAT shows that “on average, participants were significantly faster at pairing litigator-related traits with White faces compared to Asian faces”. The explicit measures, on the other hand, found little bias except for the questions about how “most Americans” feel, which showed that participants believed that “most Americans” held the very biases which they themselves were denying they held.

The test also showed correlations between IAT scores and how the identical litigators were rated. The more participants implicitly envisioned the ideal litigator to be White, the more competent, warm, and hireable they thought the White deposing litigator was. Meanwhile, “evaluation of the Asian American litigator was significantly correlated with [subjects’] explicit stereotypes.” In other words, “people’s evaluations of the White litigator’s performance was most strongly related to their implicit stereotypes of who they envisioned as the ideal litigator, whereas their evaluations of the Asian litigator’s performance was most strongly related to their explicit stereotypes about the ideal litigator. Both explicit and implicit bias may therefore determine our assessment of someone’s job performance.

The research evaluating the racial and gendered components of employer evaluation of efficacy is vast. In an article entitled Discrimination-Reducing Measures at the Relational Level, Tristin Green and Alexandra Kalev provide a thoughtful evaluation of this social science research. According to a wide variety of researchers, race and gender operate in myriad ways to depress the performance evaluation of underrepresented groups. It has long been recognized by social scientists that when members of minorities and women are substantially underrepresented in the work place, they are likely to be judged according to stereotypes. These stereotypes then often have the effect of creating minimized expectations of women and people of color – which leads to differential behavior toward these employees, which can then result in women and employees of color being perceived as performing less well.

The research provides us reason to be optimistic, however. As Green and Kalev discuss, “priming those in power with egalitarian values can lead them to pay closer attention to information that contradicts stereotypes of outgroup members” and ensuring that women and minorities have equal status can reduce prejudice considerably.