Examining the Facts on Implicit Bias

Richard R. Johnson, Ph.D.

May, 2017

A number of sources have claimed that public employees are influenced by implicit biases. The U.S. Department of Justice, the Police Executive Research Forum, and the President's Task Force on 21st Century Policing, for example, have suggested that law enforcement officers hold unconscious, implicit biases against people of color.[1] **It has been argued that these implicit biases cause police officers to enforce the law in ways that discriminate against members of racial minority groups.** Similar claims have been made against prosecutors, judges, and probation officers as an explanation for the disproportionate representation of racial minorities in our prisons and jails. Allegations have also been levelled against teachers and school administrators, suggesting that they treat white students preferentially over minority students, and that they do so as a result of these same unconscious, implicit biases.[2]

One of the remedies often suggested to address implicit bias is some form of implicit bias training.[3] Are these claims supported by the available evidence? The purpose of this brief is the factual examination of the empirical evidence surrounding the concept of implicit bias, implicit bias tests, and the relationship between implicit bias test scores and actual discriminatory behavior.

What is Implicit Bias?

In recent years the concept of implicit bias has received a great deal of attention in the United States. Implicit bias is an idea suggesting that, regardless of our conscious thoughts and feelings, we each hold biased judgements in our subconscious against people that are different. For example, it has been argued that Caucasian people who make a concerted effort to avoid discriminating against African-Americans still hold untrue racist stereotypes and opinions about African-Americans in their subconscious minds.[4] It has even been asserted that individuals often hold negative implicit bias attitudes toward members of their own racial or gender group. In other words, African-American teachers may be more punitive toward African-American students because society has imbedded into their subconscious false stereotypes about African-Americans as poorly performing students.[5]

It has been asserted that these hidden, unconscious biases cause individuals to act in discriminatory ways toward others, even though individuals do not consciously intend to do so. Advocates of the concept of implicit bias suggest that these unconscious biases result in many instances of discrimination against women and members of racial minority groups. These instances of discrimination include hiring, promotion, and assignment discrimination in the workplace, grade and punishment discrimination within schools, and diagnostic and treatment decisions within hospitals and doctors' offices.[6] Accusations have been made that implicit bias is affecting the decisions of those who work within the criminal justice system. Police officers, prosecutors, judges, corrections officers, and probation / parole officers have been accused of making decisions biased against African-Americans due to implicit biases.[7]

The concept of implicit bias, first developed by sociologists during the 1960s, has led to the creation of psychological tests in the 1990s that purport to measure one's unconscious, hidden biases.[8] Training workshops have sprouted up in the 2010s that are designed to help individuals discover and confront their implicit biases.[9]

How do Implicit Bias Tests Work?

The first step in addressing implicit bias is to determine if one suffers from it. As mentioned above, in the 1990s some psychologists began attempting to design tests that could measure one's implicit bias. Today there are a variety of tests that purport to measure implicit bias with regard to a number of biases, such as racist biases, sexist biases, homophobic biases, socioeconomic biases, and biases against persons with disabilities.[10] Because there are a number of tests, created by various researchers, designed to test for different types of biases, it is difficult to address every one of these tests. This report, therefore, will focus primarily on the most commonly used test, known as the Implied Association Test (IAT). The IAT has variations designed to test for different types of biases (i.e., racial versus gender biases), but all of the IAT tests use a similar methodology.[11]

An IAT generally operates in the following manner. First, the test-taker is required to complete a sorting task on a computer. The test-taker is presented with one category on the right side of the screen, and a second category on the left side of the screen. Words will appear in the center of the screen and the test-taker must assign the word to one of the two categories by hitting a corresponding key on either the right or left side of the computer's keyboard. For example, the categories into which things must be sorted might be "good" and "bad." The test-taker is given meaning-laden words such as "despicable," "ethical," or "evil" that must be assigned to one category or the other ("good" or "bad"). The test-taker is encouraged to answer as quickly as possible and respond to the first impulse that comes to mind. Next, the test-taker completes another categorization exercise, this time categorizing photos of faces of whites and blacks into categories labelled as "black" or "white." The test-taker must then sort the photos into the two categories, again using the computer keys and with encouragement to continue to answer swiftly.

As the test continues, the test then begins to use categories that pair the previous categories encountered. For example, the test-taker may have to sort into the categories "black / good" and "white / bad." When a photo of an African-American face appears, the test-taker should assign the photo to the "black / good" category because that is where the black face belongs (i.e., because of

the "black" category, not the "good" category). When the test-taker encounters a word like "wicked," it should be assigned to the category "white / bad" because that is where words denoting "bad" belong (not because of any association with the category "white"). The test-taker must rapidly assign a quick succession of both faces and good / bad attribute words. As the test goes on, the categories will switch pairings (i.e., "white / bad" will become "white / good" and vice versa) and the category pairings will also switch sides of the screen as the test-taker has to quickly keep categorizing names and attribute words appropriately for several minutes.

What the test is measuring is how often the test-taker associates the bias category – racial bias against African-Americans in this example – with the two different attribute categories. For example, does the test-taker, while rapidly trying to categorize words into categories that quickly change titles and positions on the screen, assign more of the negative attribute words to one race group than the other? In our example here, we would be measuring if the test-taker assigns (correctly or incorrectly) more words associated with "bad" to the "black" category and more words associated with "good" to the "white" category. The degree these words are not equally assigned to both race categories is supposed to indicate the degree of racial implicit bias the test-taker holds.[12]

As this is how implicit bias is purportedly measured, **obvious flaws appear to be evident in the detection of implicit biases. First, the fact that in study after study, almost all individuals test positive for some level of implicit bias, even against their own racial and gender group, has caused many researchers to suggest that this is evidence these tests lack validity. Women who consider themselves to be very liberal and strongly identify with feminist ideals still test positive for implicit sexist biases against women. Likewise, African-American individuals who identify themselves to be very liberal and consider themselves to be conscious of race issues can also test positive for biases against blacks.** This has led many scientists to question the validity of these tests.[13] Imagine someone developed a test to detect stomach cancer. What if, during clinical trials with people with and without cancer, 100% of those who took the test tested positive for stomach cancer? The researchers involved would question the test's validity. The same thinking should apply here.

Another flaw is that there are no middle-ground options to the test-taker. The test-taker may think neither whites nor blacks should be associated with words like "good" or "bad," yet the test-taker is forced to make a categorization one way or the other. In cases like this, the test does not measure the test-taker's real attitude, only his or her forced response to contrived categorization exercises designed to force the test-taker to answer in a biased manner toward one group or another. Perhaps this explains why people who no one would ever suspect of being biased – especially against their own group – still can test high for implicit bias.

A third issue with IAT test methods is that while they claim to be measuring stereotypes about certain groups, they are sometimes actually measuring statistical realities. On some racial bias IAT tests, for example, other categorizations are used. These tests require the test-taker to categorize whites and blacks by attributes that include wealth or crime involvement. In these cases, is a person truly racist if he or she, under the stress of time and moving categories, categorizes "poor" as a black trait and "rich" as a white trait? After all, according to the U.S. Census Bureau, 27% of African-Americans live in poverty while only 10% of non-Hispanic whites

live in poverty.[14] African-Americans, therefore, are 170% more likely than are whites to live in poverty. This is why the government and nonprofit organizations have poverty relief programs dedicated to African-American communities. **When discussing generalities, African-Americans *are generally* less financially wealthy than are whites. Again, this is statistical reality rather than an implicit bias. Also remember that this is a forced choice, and the test-taker must choose one of the two categories with no "neither" option.**

Implicit Bias Tests and Actual Biased Behavior

If IATs actually measure implicit biases, and these unconscious biases affect our behaviors, then there should be a direct correlation between a person's implicit bias test score and the degree of bias in that person's behaviors. Numerous empirical research studies have now been conducted to examine whether this correlation actually exists.

Many individual studies have examined whether the IAT, or other implicit bias tests, can predict actual biased behaviors. There have been numerous individual studies, examining different types of biases, using varying sample sizes, and various measures of biased behavior. These studies have produced a mixture of findings that have often conflicted, in part because the individual differences between tests in how they were conducted differed so markedly. Several teams of researchers, however, have tried to decipher if any trends have appeared across these studies.

In 2009, the research team that originally created the IAT tests conducted a review of the existing studies at the time. After reviewing 122 studies of various IAT test types (racial test, gender test, etc.), this team concluded that there was evidence implicit bias test scores predicted peoples' biased judgements and behaviors, though the influence was small.[15]

Additional major reviews in 2009 and 2013 conducted by large, neutral, independent teams of researchers from numerous major universities across the U.S. have found different results. These teams of scientists came from such prestigious institutions as the University of California at Berkeley, University of Pennsylvania, University of Virginia, University of Connecticut, Texas A&M University, New York University, Rice University, City University of New York, and Florida International University. Because there were so few studies examining other types of bias, these researchers only examined studies that evaluated the racism IAT and excluded studies that examined other types of bias, such gender bias, or sexual-orientation biases.[16]

These teams examined 46 published studies that involved 5,600 total persons tested. All of these studies involved college students who were given the racial bias IAT test before participating in laboratory experiments designed to measure biased behaviors. The experiments involved such activities as rating the academic ability of students of various races, engaging in board games that required varying levels of cooperation with partners of different races, and conducting mock job interviews of job applicants of various races. **The examination of all 46 studies in existence at the time of the investigation revealed that racial bias IAT test scores showed no significant relationship with biased behaviors observed during these experiments. Study after study showed either no association between IAT scores and biased behavior, or only a very weak association**.[17]

The research teams also noted that only some of the studies tried to control for overt racist attitudes when measuring implicit bias. In other words, did people that openly expressed racist attitudes also score high in implicit bias? The research team found that people who openly espouse racist beliefs generally were no more or less likely to have a high implicit bias score, but those who were openly racist did demonstrate more racially biased behaviors during the experiments. After removing those who were openly racist, the researchers found that the implicit bias test scores of the test subjects that remained had no correlation with racially biased behaviors in the experiments. In other words, **openly biased attitudes seemed to predict racially biased behaviors rather than individuals with some hidden, deeply unconscious psychological bias**.[18]

These findings were published in many prominent psychology research journals and additional studies are continually produced today that question the validity of implicitly biased behaviors. Here are just a few criminal justice system examples.

One study involved a sample of law students who were given the sexism IAT as well as a test to measure their openly sexist attitudes toward women. The sample of law students was almost equally divided between men and women. After completing their IAT, these future lawyers were then put through several tasks, such as judging the attributes of several persons being considered for appointment to positions as judges, and proposing budget cuts to the university by eliminating organizations, services, and programs that might impact men and women differently.[19]

This study found that most of the law students, regardless of sex, showed some level of implicit bias against women on the IAT test. No relationship was found between the implicit bias scores and the scores on the openly sexist test. This shows that even some who are openly sexist score low on implicit sexist bias, and vice versa. The participants' implicit bias test scores were also found to be completely unrelated to any gender bias they displayed in the two exercises they were given – rating the judicial appointees and making the budget cuts. In fact, very little gender bias was shown by any of the participants in these two tasks.[20]

Another study involved a sample of 80 experienced police officers who completed the IAT test for implicit racism. After completing the implicit bias test, each officer engaged in 24 very realistic use of lethal force training scenarios over the course of four days. The officers encountered "shoot / don't shoot" video scenarios in a firearms training simulator (FATS) machine that required the officers to decide, under stress, whether or not it was legally appropriate to shoot a potential criminal suspect visually displayed in an interactive video scenario. If the suspect's behavior required the officer to use lethal force, the officer was expected to draw and fire his or her weapon at the suspect. Computerized laser technology measured how quickly the officer drew the weapon, fired the weapon, and whether or not the officer's aim would have resulted in a bullet accurately hitting the suspect. In each of these high-stress scenarios the race of the suspect was manipulated so that in some scenarios the suspect was black, and in others the suspect was white. The circumstances of the scenarios were also altered so that in some scenarios lethal force would have been justified, but not in others.[21]

The researchers measured if the officers were more likely to shoot the suspect in error (when not legally justified to do so) if the suspect was black as opposed to white. The researchers also looked for race differences between white and black suspects in how quickly the officers drew and fired

their weapons, and how accurately they shot. **The findings revealed that the officers were generally slower to shoot armed black suspects posing a lethal threat than they were to shoot armed white suspects, suggesting a bias toward shooting whites rather than blacks. The officers rarely shot a suspect when not legally justified to do so, but when errors did occur, whites were more likely to be shot incorrectly.** Shooting accuracy did not appear to differ by suspect race. **Importantly, the officers' implicit bias test scores were completely unrelated to their shooting decision behaviors.** Neither a high, nor a low, implicit bias score resulted in biased shooting behavior against either black or white suspects.[22]

Implicit Bias Training

Supporters of the implicit bias idea have asserted that training is needed to reveal to individuals their implicit biases, and help these individuals to overcome their hidden biases when they act within the workplace. Implicit bias training seminars are widespread and government entities and businesses across the nation are requiring their employees to undergo implicit bias training. For example, numerous school districts are requiring staff to complete this training, as are law enforcement agencies, courts, and hospitals.[23]

At present there are no published studies that have evaluated the ability of these implicit bias training courses to reduce implicit bias attitudes or behaviors.[24] This is surprising since all one needs to do is have the participants complete an IAT test before the training and then repeat the test after the training to see if the participants' implicit bias test scores when down. Unfortunately, because no evidence exists to determine the effectiveness of these implicit bias training courses, we can only speculate as to their effectiveness by comparing this training to the effectiveness of racial diversity training.

Even though racial diversity training has been conducted for decades, there is little published research on this topic with regards to whether or not this training has any effect on attitudes or behaviors. **Reviews of the 30 or so existing studies have evaluated the influence of multicultural education on racial attitudes among grade school students, college students, and police officers. These studies have consistently revealed no lasting influence on participants' racial attitudes**, especially attitudes toward African-Americans.[25] As it is safe to say that there is no evidence that traditional multicultural or racial diversity training has any significant positive influence on people's attitudes or behavior, how much less likely is implicit bias training to have an effect on attitudes and behaviors?

Proponents of the implicit bias concept suggest that implicit biases are deeply ingrained in our unconscious mind. Assuming this was true, how could attending a training session be expected to have any impact on one's subconscious thoughts and motives? It would seem more plausible that such a condition could only be corrected through many months or years of intensive therapy. Currently, the only psychological therapies used to address unconscious beliefs involve hypnosis and sedative-hypnotic drugs.[26]

Therefore, if implicit biases actually exist, it is doubtful that mandating training is likely to impact it. If, on the other hand, implicit bias does not really exist, such implicit bias training

only serves to insult and harm the attendees by trying to convince them they possess biases that they do not.

Conclusions

The empirical research evidence reviewed here leads to four major conclusions. First, implicit bias test scores may lack validity, simply by their construction alone. The forced choice nature of the categorization tasks they require compel people to associate politically charged words with one group or another, and do not permit the test-taker to suggest the words are equally associated with both groups. Test responses are judged to be prejudicial even when they are statistically more accurate generalities of the group described (i.e., whites *generally* wealthier than blacks). Even people who outwardly show no signs of bias can score high in bias on these tests, and **research has repeatedly revealed no association between implicit bias test scores and openly biased attitudes**. These facts call the validity of implicit bias tests into question.

The second major conclusion is that there is no significant link between implicit bias test scores and actual biased behaviors. No matter whether one is examining games and role-playing exercises between undergraduate college students in psychology labs, or realistic lethal force decision-making scenarios with experienced police officers, **time and again implicit bias test scores fail to predict any actual biased behaviors.**

A third conclusion is that implicit bias, if it exists, is unlikely to be corrected or transformed by any interventions that short-term group training can offer. If such biases exist at the deeply subconscious level, the social or psychological processes that created them must have been significant and accumulated over some length of time. As a result, it seems irrational to believe that simply taking a training course would reverse their influences. If these biases operate at the deeply unconscious level, it would seem unreasonable to believe that simply being aware that these biases exist would be enough to overcome them. **There seems to be no available evidence indicating that implicit biases are corrected or transformed by training.**

The final conclusion that can be reached from this report is the clear evidence that holding explicit, open biases *do* result in biased behaviors. If an individual holds openly racist views against African-Americans, that person is likely to act in ways that discriminate against African-Americans. If a person is openly sexist and hostile towards women, then that person is very likely to act prejudicially against women. **While the research supporting implicit bias and its alleged effects is very weak, there is little doubt about the effects of overt, openly biased attitudes on behavior.**

The public policy implications of these conclusions appear clear. First, it would be counter-intuitive for organizations such as schools, hospitals, law enforcement agencies, courts, and businesses to utilize finite training resources to address a perceived hidden bias problem for which there is little credible evidence. Perhaps the money and time spent on training and testing regarding implicit biases could be put to better use elsewhere. Second, it would seem counterproductive to try to improve race and gender relations by falsely accusing employees of being subconscious racists or chauvinists. These inflammatory assertions are being made without significant proof.

The final clear policy implication is that those who are openly biased toward other groups should be prevented from holding positions of authority. People who are openly racist should never hold life-and-death power over other people, such as in the role of a police officer or judge. People who are openly sexist should never be given the opportunity to decide who gets hired into an organization or promoted within the organization. **The evidence suggests that people who are openly biased in word are openly biased in deed. The evidence also suggests that those who do not hold consciously biased views do not have to worry that they are actually, deep down in their subconscious, bigots.**

References

[1] Edwards, J. (June 27, 2016). Justice Department mandates implicit bias training for agents, lawyers. *Reuters.* Downloaded from: http://www.reuters.com/article/us-usa-justice-bias-exclusive-idUSKCN0ZD251; Fridell, L. (June, 2014). Psychological research has changed how we approach the issue of biased policing. *Subject to Debate,* 4-5; President's Task Force on 21st Century Policing (2015). *President's Task Force on 21st Century Policing, Final Report*. Washington, DC: Office of the U.S. President.

[2] Staats, C. (2016). Understanding implicit bias: what educators should know. *American Educator, 39(4)*, 29-43.

[3] Parascandola, R. (July 15, 2016). NYPD to develop training curriculum based on noted academic Dr. Phillip Goff's analysis of implicit bias. *New York Daily News*. Downloaded from: http://www.nydailynews.com/new-york/exclusive-nypd-develop-implicit-bias-training-curriculum-article-1.2713332

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

[5] Anderson, M. D. (September 28, 2016). Even black preschool teachers are biased. *The Atlantic*. Downloaded from: http://www.theatlantic.com/education/archive/2016/09/the-high-standard-set-by-black-teachers-for-black-students/501989/

[6] Jost, J. T., Rudman, L. A., Blair, I. V., Camey, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: a refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior, 29*, 39-69.

[7] Edwards, J. (June, 2016).

[8] Greenwald & Banaji (1995); Jost et al. (2009).

[9] Flannery, M. E. (September 9, 2015). When implicit bias shapes teacher expectations. *NEA Today*, Downloaded from: http://neatoday.org/2015/09/09/when-implicit-bias-shapes-teacher-expectations/; Fridell, L. (2016). *Fair and Impartial Policing: A Science-Based Approach*. New York, NY: Springer Publishing.

[10] Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: scientific foundations. *California Law Review, 94(4)*, 945-967.

[11] Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464–1480.

[12] Greenwald, McGhee, & Schwartz (1998); Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review (pp. 265–292). In J. A. Bargh (Ed.), *Automatic Processes in Social Thinking and Behavior*. New York, NY: Psychology Press.

[13] Blanton, H., & Jaccard, J. (2015). Not so fast: ten challenges to importing implicit attitude measures to media psychology. *Media Psychology, 18(3)*, 338-369; Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. (2015). Using the IAT to predict ethnic and racial discrimination: small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology Studies, 108(4)*, 562-571.

[14] U.S. Census Bureau (2016). *Income and Poverty in the United States: 2015*. Washington, DC: U.S. Census Bureau.

[15] Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test II: Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97(1)*, 17-41.

[16] Blanton, H., & Jaccard, J. (2015). Not so fast: ten challenges to importing implicit attitude measures to media psychology. *Media Psychology, 18(3)*, 338-369; Blanton, H., Jaccard, J., & Burrows, C. N. (2015). Implications of the implicit association test D-transformation for psychological assessment. *Assessment, 22(4)*, 429-440; Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94(3)*, 567-582; Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: strange and stranger. *Industrial and Organizational Psychology, 1(4)*, 379-392; Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. (2013). Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology Studies, 105(2)*, 171-192; Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. (2015). Using the IAT to predict ethnic and racial discrimination: small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology Studies, 108(4)*, 562-571.

[17] Ibid.

[18] Ibid.

[19] Levinson, J. D., & Young, D. (2010). Implicit gender bias in the legal profession: an empirical study. *Duke Journal of Gender Law and Policy, 18(1)*, 1-44.

[20] Ibid.

[21] James, L., James, S. M., & Vila, B. J. (2016). The reverse racism effect: Are cops more hesitant to shoot black than white suspects? *Criminology and Public Policy, 15(2)*, 457-479; James, L., Klinger, D. A., & Villa, B. J. (2014). Racial and ethnic bias in decisions to shoot seen through a stronger lens: experimental results from high-fidelity laboratory simulations. *Journal of Experimental Criminology, 10(3)*, 323-340.

[22] Ibid.

[23] Edwards, (2016); Fridell, L. (2016). *Fair and Impartial Policing: A Science-Based Approach*. New York, NY: Springer Publishing; President's Task Force on 21st Century Policing (2015); Staats, (2016).

[24] See Zarya, V. (November 10, 2015). I failed this test on racism and sexism – and so will you. *Fortune*. Downloaded from: http://fortune.com/2015/11/10/test-racism-sexism-unconscious-bias/

[25] Bigler, R. C. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues, 55(4)*, 687-705; Paluck, E. L. (2006). Diversity training and intergroup contact: a call for action research. *Journal of Social Issues, 62(3)*, 577-595; Paluck, E. L., & Green, D. P. (2009). Prejudice reduction What works?: a review and assessment of research and practice. *Annual Review of Psychology, 60*, 339-367; Pendry, L. F., Driscoll, D. M., & Field, S. (2007). Diversity training: putting theory into practice. *Journal of Occupational and Organizational Psychology, 80(1)*, 27-50; Stewart, T. L., LaDuke, J. R., Bracht, C., Sweet, B., & Gamarel, K. E. (2003). Do the 'eyes" have it? A program evaluation of Jane Elliott's "blue eyes / brown eyes" diversity training exercise. *Journal of Applied Social Psychology, 33(9)*, 1898-1921.

[26] Garson, L. (2006). *Surviving Babylon: A Journey through Repressed Memories of Sexual Abuse*. New York, NY: Alexander Griffin Co.; Spiegel, D. (1997). *Repressed Memories*. Washington, DC: American Psychiatric Press.