Implicit Discrimination

By MARIANNE BERTRAND, DOLLY CHUGH, AND SENDHIL MULLAINATHAN*

What drives people to discriminate? Economists focus on two main reasons: "taste-based" and "statistical" discrimination. Under both models, individuals *consciously* discriminate, either for a variety of personal reasons or because group membership provides information about a relevant characteristic, such as productivity. Motivated by a growing body of psychological evidence, we put forward a third interpretation: implicit discrimination. Sometimes, we argue, discrimination may be *unintentional* and outside of the discriminator's awareness.

I. Psychology of Implicit Attitudes

Most modern social psychologists believe that attitudes occur in both implicit and explicit modes, suggesting that people can think, feel, and behave in ways that oppose their explicitly expressed views, and even, explicitly known self-interests.[1] The preferences and beliefs that economists typically describe as an individual's "attitudes" are what psychologists would specify as "explicit attitudes," which may or may not align with the same individual's "implicit attitudes," defined as *unconscious* mental associations between a target (such as an African-American) and a given attribute.

One of the most important recent research insights is that implicit attitudes can be measured. A widely used measure of implicit mental processes is the Implicit Association Test (IAT) (Anthony G. Greenwald et al., 1998). The IAT relies on test-takers' speed of response to represent the strength of their unconscious mental associations.[2] IATs are used to measure a wide range of implicit attitudes about social groups, products, or self-identity. We illustrate this with a race IAT.

The race IAT is typically taken on a computer. The test-taker must quickly categorize words and pictures of faces that appear in the center of the screen. Faces are to be categorized as African-American or white and words (such as happiness or tragedy) as good or bad. Pairs of categories appear on either side of the screen. If the stimulus belongs to categories on the right (left), the test-taker hits a key on the right (left) side of the keyboard. Each test-taker completes two versions of the task, categorizing as many as 60 different stimuli. In one, the "compatible" version, the two categories on one side are paired according to a stereotype, such as "African-American" with "bad" in one corner, and "White" with "good" in the other corner. In the "incompatible" version, the categories are paired counter-stereotypically, such as "African-American" with "good," and "white" with "bad." The key insight of the race IAT is that an implicit bias against African-Americans shows up as a response time differential. Most people respond more quickly in the compatible pairing, when African-American is paired with bad rather than good, demonstrating a stronger mental response.

† *Discussants:* Dan Black, Syracuse University; David Neumark, Public Policy Institute of California and Michigan State University; Shelly Lundberg, University of Washington; Kerwin Charles, University of Michigan.

* Bertrand: Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60603; Chugh: Harvard Business School, Soldiers Field Road, Boston, MA 02163; Mullainathan: Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138. This paper is based on work supported by the National Science Foundation under grant no. 0351184.

[1] Due to space constraints, we omit many references. See Bertrand et al. (2005) for full references to the relevant papers.

[2] A demonstration of the test is available online: ⟨http://implicit.harvard.edu⟩.

Because people may misrepresent their explicit attitudes, perhaps the IAT is simply a less "fakable" measure. However, recent neuroscientific studies demonstrate that conscious processing activates different regions in the brain than does unconscious processing, thus these are distinctive mental processes. One study showed greater brain activity associated with control and regulation when supraliminally processing black faces, in contrast with greater brain activity associated with emotion and fear when subliminally processing black faces. Another showed a correlation between the IAT and amygdala activation (fear response) in response to black faces. In addition, the divergence of implicit and explicit attitudes is not limited to socially sensitive domains. For example, the social demands to conceal one's preferences about a Mac versus PC computer, or Coke versus Pepsi seem minimal. Yet, implicit and explicit attitudes in these domains are imperfectly correlated, with both having predictive power.

Can implicit attitudes influence behavior in meaningful ways? Evidence to date suggests yes. A meta-analysis of 61 studies found an average correlation of 0.27 between the IAT and outcome measures such as judgments, choices, and physiological responses. Most importantly, the IAT outperformed explicit attitude measures for less-controllable behavioral outcomes. In one study, white participants interacted with both a white and African-American experimenter, and also took the IAT. Participants' implicit attitudes favoring whites predicted more smiling, speaking time, extemporaneous social comments, and general friendliness, as well as fewer speech errors and speech hesitation, toward the white experimenter.

These findings suggest that controllability may be an important behavioral dimension. But could any relevant economic behavior, such as a hiring decision, truly be characterized as "hard-to-control"? In fact, social psychologists argue that even theoretically controllable behaviors may operate with greater automaticity under certain situational conditions. Chugh (2004) described the "messy, pressured, and distracting" conditions of managerial work as conducive to implicit mental processes. Time pressure and stress are two situational influences likely to first generate an acceleration of the mental process, and then an attempt to reduce the amount of information needing processing. This type of "cognitive load," also occurs in the form of conflicting yet simultaneous task demands and excessive attentional demands.

In addition, social psychologists argue that many seemingly controllable behaviors may be prone to implicit attitudes under conditions of ambiguity, and have demonstrated that implicit discrimination is more likely to occur in situations where multiple, non-racist explanations for the behavior might exist. Thus, some conditions under which implicit attitudes may arise are threefold: inattentiveness to task, time pressure or other cognitive load, and ambiguity.

II. Can Implicit Attitudes Be "Manipulated"?

One intriguing feature of implicit attitudes is their potential manipulability. In one study, white participants were told they would be working with a black individual, who would either be their subordinate or their superior. Those anticipating a black superior showed more positive implicit attitudes toward blacks than those anticipating a black subordinate, suggesting that positive and powerful black exemplars are important cues. In another, exposure to photographs of admired African-Americans (e.g., Bill Cosby) led to a decrease in anti-black implicit attitudes, an effect that persisted for 24 hours. In another, reducing attention to race cues (e.g., by increasing attention required by the task) moderated implicit attitudes. This work certainly does not imply that implicit attitudes can be reversed with simple manipulations of the situation or task. However, the work suggests malleability in implicit attitudes and associated behaviors.

III. Interpreting Existing Audit Studies in the Light of Implicit Discrimination

Obviously, implicit attitudes cannot explain all forms of racial discrimination. Explicit discrimination in employment ads prior to the Civil Rights Act of 1964 had little to do with implicit attitudes. However, we find it reasonable to hypothesize that several other documented forms of differential treatments may, in part, reflect such implicit attitudes.

The Bertrand and Mullainathan (2004) résumé task, for example, theoretically satisfies

several criteria thought to be important for implicit discrimination to arise. First, the task is typically performed under important time pressure, as the screeners have to make their way through a thick pile of résumés, often juggling this task with multiple other administrative loads. The task is also involves considerable ambiguity: in the search for a "good" job applicant, there is no such thing as a simple formula to be followed to determine which candidates are above the "fit line." Also, the typical task is a nonverbal automatic process consisting in placing a given résumé either on the "yes" pile or on the "no" pile, with little commentary on each résumé.

Several other field experiments may fit the implicit discrimination model. Consider Ian Ayres et al.'s (2004) finding of African-American cab drivers receiving lower tips than white cab drivers. A tipping decision is often made quickly, just as the passenger is stepping out of the cab, and when the passenger's mind is preoccupied with an upcoming destination or event. Finally, ambiguity exists in how to interpret subtle cues about friendliness and honesty.

Bargaining is another relevant context, as in John List's (2004) study of discrimination in the sports-card market. When a prospective buyer expresses interest in a card, the seller makes a quick first offer. Very often, this first offer is made as the seller's attention is split between the current buyer and other prospective buyers nearby.

Also, consider the housing audit studies documenting differential treatment of equally qualified African-American and white home buyers in realtors' showing of additional units, both in terms of numbers and quality (see e.g., Jan Ondrich et al., 2003). The realtor faces a subtle, complex, and ambiguous task in forecasting a client's idiosyncratic tastes.

A police officer's decision of whether or not to shoot a potentially armed target is taken in an ambiguous split second. Joshua Correll et al. (2002) used a videogame to show that subjects were quicker at deciding not to shoot an unarmed white target versus an unarmed black target, even though both targets were armed at equal rates in the context of this game. Most interestingly, the authors showed that this difference was not related to cross-subjects differences in explicit racial prejudice.

IV. Testing for Implicit Discrimination

Hence, implicit discrimination could potentially explain some economic phenomena, with sufficient testing. We suggest several potential directions for future research.

A first approach would be to perform more correlation exercises in the field between economic behavior and IAT. One could contact the realtors after a fair-housing audit took place and ask them to take an IAT, or contact sports-card traders studied by List (2004). Alternatively, with some creativity, one might integrate a field element within a lab study. For example, if taxicabs pick up subjects to bring to the lab for an IAT, one could correlate subjects' IAT scores with their tipping behavior.

Second, one could perform additional tests by empirically varying situational factors shown to be important for implicit attitudes to affect behavior. For example, one could schedule an appointment with a realtor either when s/he is quite busy or less busy. Or one could vary the level of ambiguity of the realtor's task with a more-specific or less-specific description of the client's desired home.

One could also reduce attention to the social cues in the context of the résumé study by modifying the location of the names on résumés. Bertrand and Mullainathan (with Abhijit Banerjee) are currently carrying out such a manipulation in India in the context of caste-based discrimination. In India, it is possible for a given individual to have a caste-neutral name but for his or her father to have a lower-caste name. It is also common for an individual to report the father's name at the *bottom* of the resume. One can therefore compare callback rates for lower-caste people whose caste affiliation is communicated through their names versus through their father's names.

Another testing possibility is to attempt to mimic natural situations in the laboratory itself. We have started exploring this possibility in the context of the résumé study. Specifically, we recruited 115 subjects for a study on information-processing and attention. The task was to screen 50 résumés for a company filling an administrative assistant position (job description provided). Their task was to select the 15 best candidates. Each participant received a unique set of résumés

in that, following Bertrand and Mullainathan (2004), each résumé was randomly assigned either a white-sounding or African-American sounding first name. After completing this task, the participants took several IATs, answered explicit attitude measures about African-Americans, and completed a debriefing survey ("how rushed did you feel ... ?"). Anonymity on all measures was fully guaranteed to all participants.

While our pilot testing findings are preliminary, some encouraging results have emerged. First, participants who reported feeling rushed picked a significantly lower fraction of résumés with African-American names. We also found a negative correlation between the number of African-American résumés selected by a given subject and that subject's implicit attitude about intelligence in blacks and whites (where negative scores indicate an association between African-American and dumb). Most interestingly, this negative correlation was concentrated among those subjects who *ex post* reported feeling most rushed during the task. In contrast, we found no apparent correlation between the number of African-American résumés picked and the self-reported explicit attitudes towards African-Americans.

Obviously, such a lab exercise lacks external validity and faces implementation problems. In this regard, the subjects' background (mostly undergrads) and the difficulty of providing naturalistic incentives may explain one major issue with our pilot study so far: we did not find discrimination, on average, in the lab and only those subjects who felt rushed picked a lower than base-rate fraction of African-American résumés. In the future, we hope to implement a similar exercise within a firm.

Also, once the design is perfected, we could test de-biasing remedies that emerge naturally from the psychological evidence. First, and most obvious, one might simply inform human-resource managers about the existence of the implicit bias. Second, small changes in the situational context of résumé screening could have potential large positive effects. Simply leaving more time to the screeners to assess the merit of each résumé may limit the role for unconscious responses while performing this task. Also, having an African-American person in the interview room, or even in mind, may operate as a positive exemplar (not a monitor) which could mute the importance of unconscious reactions. Also, a more structured review process that draws attention to the task cues rather than social cues (such as highlighting the positive and negative aspects of each résumé, or evaluation along highly specific job criteria, rather than a general "fit" comparison to a broad job description).

V. Conclusion

However we test for it, implicit discrimination is not useful simply as a subtle alternative interpretation. If it is a powerful driver of discriminatory behavior, it should reshape the way we understand discrimination and alter our available spectrum of remedies. A key differential feature of potential remedies to implicit discrimination is that they could limit the amount of discrimination without forcing agents to take decisions against their will. In fact, because people may be engaging in injurious behavior without realizing it, the remedies may bring their decisions closer in line with what they (explicitly) think or favor for their organization. Another important feature of these remedies is that, unlike most affirmative-action policies, they can be implemented at low cost and without making race salient, greatly increasing political feasibility.

REFERENCES

Ayres, Ian; Vars, Fred and Zakariya, Nasser. "Racial Disparities in Taxicab Tipping." Working paper, Yale Law School, 2004.

Bertrand, Marianne; Chugh, Dolly and Mullainathan, Sendhil. "Implicit Discrimination." Working paper, Graduate School of Business, University of Chicago, 2005.

Bertrand, Marianne and Mullainathan, Sendhil. "Are Emily and Greg More Employable Than Lakisha and Jamal?" *American Economic Review*, 2004, *94*(4), pp. 991–1013.

Chugh, Dolly. "Why Milliseconds Matter: Societal and Managerial Implications of Implicit Social Cognition." *Social Justice Research*, 2004, *17*(2), pp. 203–22.

Correll, Joshua; Park, Bernadette; Judd, Charles M. and Wittenbrink, Bernd. "The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening

Individuals." *Journal of Personality and Social Psychology*, 2002, *83*(6), pp. 1314–29.

Greenwald, Anthony G.; McGhee, Debbie E. and Schwartz, Jordan L. K. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology*, 1998, *74*(6), pp. 1464–80.

List, John. "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field." *Quarterly Journal of Economics*, 2004, *119*(1), pp. 49–89.

Ondrich, Jan; Ross, Stephen and Yinger, John. "Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Available Houses from Black Customers?" *Review of Economics and Statistics*, 2003, *85*(4), pp. 854–73.

References

Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

Marianne Bertrand; Sendhil Mullainathan

The American Economic Review, Vol. 94, No. 4. (Sep., 2004), pp. 991-1013.

Stable URL:

http://links.jstor.org/sici?sici=0002-8282%28200409%2994%3A4%3C991%3AAEAGME%3E2.0.CO%3B2-H