Implicit Bias, Law and Policy: Annotated Bibliography
July 2011

Prepared by Jasmine Jeffers, PRRAC Intern

Overview:

A primer on implicit bias from UCLA Law Professor Jerry Kang:
http://jerrykang.net/2011/03/13/getting-up-to-speed-on-implicit-bias/

Law Reviews:


Summary: Bagenstos produces a response to current antidiscrimination norms and their critics, arguing that antidiscrimination activists must divide their efforts between establishing the scientific validity of implicit bias and determining what role its existence should play in anti-discrimination litigation. Through an analysis of Mitchell & Tetlock’s 2006 study, he develops the argument that many anti-implicit bias arguments are not an inherent denouncement of the concept’s scientific standing, but stand to provoke valuable discussion about the normative basis of antidiscrimination law.


Summary: A critique of the Implicit Association Test (IAT), a computer-administrated measure of unconscious racial bias, and its implications for political discourse. The authors critique the reported conclusions of the test, questioning whether or not the findings are in fact indicators of unconscious racial bias, as opposed to concealed racial bias, and what implications, if any, the test should have for statutory or constitutional doctrine. The authors state that the unconscious bias discourse may subvert the cause of racial justice, diverting resources and attention from more immediate methods of racial reform.

Author’s Summary: This Article examines the basic legal framework established by Title VII and considers whether, and how, the broad program of research on implicit bias might fit into this scheme. We focus primarily on the admissibility of expert opinion, an issue that necessarily depends on the meaning of the law and the import of social science.


Summary: Jolls and Sunstein explore the implications of the IAT for antidiscrimination law, emphasizing its most troubling implication – that legal decisions which should be explicitly grounded in logical reasoning are being unjustly influenced by unconscious and illogical instinct. This discussion is followed by a critique of the legal field’s existing strategies for reducing the effects of implicit bias. The authors argue that most current antidiscrimination legal strategies attempt to insulate outcomes from the effects of bias – the legal field might be better served by a “debiasing” strategy, in which strategic education leads to an overall reduction in levels of racial bias.


Abstract: Once upon a time, the central civil rights questions were indisputably normative. What did “equal justice under law” require? Did it, for example, permit segregation, or was separate never equal? This is no longer the case. Today, the central civil rights questions of our time turn also on the underlying empirics. In a post–civil rights era, in what some people exuberantly embrace as post-racial, many assume that we already live in a colorblind society. Is this in fact the case? Recent findings about implicit bias from mind scientists sharply suggest otherwise. This Article summarizes the empirical evidence that rejects facile claims of perceptual, cognitive, and behavioral colorblindness. It then calls on the law to take a “behavioral realist” account of these findings, and maps systematically how it might do so in sensible, nonhysterical, and evidence-based ways. Recognizing that this call may be politically naive, the Article examines and answers three objections, sounding in “junk science” backlash, “hardwired” resignation, and “rational” justification.

Abstract: Title VII's disparate treatment model of discrimination is premised on the notion that intergroup bias is motivational in origin. This premise, in turn, is based on a number of assumptions regarding the nature of human inference and the respective roles played by cognition and motivation in social judgment and decision-making. Applying insights from cognitive psychology, Professor Krieger examines the assumptions about human inference embedded in current disparate treatment theory and questions the premise that discrimination necessarily manifests intent or motive. She suggests that a large number of biased employment decisions result not from discriminatory motivation, as current legal models presume, but from a variety of unintentional categorization-related judgment errors characterizing normal human cognitive functioning. Because of the lack of fit between the present disparate treatment model and the phenomenon it purports to represent, courts and litigants are presented with a confusing array of increasingly ill-defined and questionably premised analytical paradigms. Worse, as currently constructed, it may be exacerbating intergroup tensions and inflating both social and financial adjudication costs. Searching for solutions, Professor Krieger explores the legal and policy implications of a cognitive process approach to discrimination and equal employment opportunity and evaluates a variety of modifications to existing equal employment opportunity law.


Abstract: This Article argues that judges and jurors unknowingly propagate racism through their legal decisions because they misremember case facts in implicitly biased ways. Such an argument bridges discourse in implicit social cognition, memory studies, and legal decision-making. Social cognition research indicates that implicit racial biases are automatic, pervasive, and often operate without conscious awareness. Memory studies demonstrate that memory errors are meaningful and systematic, and are often facilitated by stereotypes. Decision-making theory teaches that memory errors can not only affect the results of individual decision-making processes, but also that group deliberations are unlikely to resolve these errors. The author argues that these three areas converge, resulting in a model that explains how implicit memory errors are made unconsciously and in racially biased ways. These unconscious and pervasive implicit memories biases consequently have the ability to affect a variety of legal outcomes.

Abstract: Recent legal scholarship challenges the default psychological assumption in antidiscrimination law that discrimination is a function of psychological processes under the conscious control of the discriminator, and replaces it with the assumption that discrimination is the result of unconscious, or implicit, psychological processes that operate automatically, beyond conscious control. This challenge is however only as persuasive as the research on which it is predicated, and we document that this research fails to satisfy key scientific tests of validity. We conclude that implicit prejudice research should be accepted as neither legislative authority nor litigation evidence until there is more: (1) rigorous investigation of the error rates of the new implicit measures of prejudice (and of how investigators balance Type I errors of false accusations against Type II errors of failing to identify prejudice); (2) thorough analysis of how well implicit measures of prejudice predict discriminatory behavior under realistic workplace conditions; (3) open debate about the societal consequences of setting thresholds of proof for calling people prejudiced so low that the vast majority of the population qualifies as prejudiced.


Abstract: Racial bias in election administration - more specifically, in the interaction between pollworkers and voters at a voting booth on election day - may be implicit, or unconscious. Indeed, the operation of a polling place may present an “optimal” setting for unconscious racial bias to occur. Pollworkers sometimes have legal discretion to decide whether or not a prospective voter gets to cast a ballot, and they operate in an environment where they may have to make quick decisions, based on little information, with few concrete incentives for accuracy, and with little opportunity to learn from their errors. Even where the letter of the law does not explicitly allow for a pollworker to exercise discretion, there is a strong possibility that unconscious bias could play a role in pollworker decision-making. Whether a poll workers’ discretion is de jure or de facto, the result may be race-based discrimination between prospective voters. This article addresses how unconscious bias may play a role in the interaction between pollworkers and prospective voters and discusses some ways in which the potential for unconscious bias to operate in America’s polling places may be mitigated.

Abstract: The disparate impact theory has long been viewed as one of the most important and controversial developments in antidiscrimination law. In this article, Professor Selmi assesses the theory's legacy and challenges much of the conventional wisdom. Professor Selmi initially charts the development of the theory, including a close look at Griggs v. Duke Power Co. and Washington v. Davis, to demonstrate that the theory arose to deal with specific instances of past discrimination rather than as a broad theory of equality. In the next section, Professor Selmi reviews the success of the theory in the courts through an empirical analysis and concludes that the theory has had a strikingly limited impact outside of the context of written employment tests and is, in fact, an extremely difficult theory on which to succeed. In the final section, Professor Selmi contends that whatever gains the disparate impact theory has produced could likely have been obtained through other means, particularly in large urban cities, and that the theory may have had the unintended effect of limiting our conception of intentional discrimination. Disparate impact theory has always been seen as beginning where intentional discrimination ends, and by pushing an expansive theory of impact we were left with a truncated theory of intentional discrimination, one that continues to turn on animus and motive. Rather than a new legal theory of discrimination, what was needed, Professor Selmi concludes, was a greater societal commitment to remediyeing inequities and the ultimate mistake behind the theory was a belief that legal theory could do the work that politics could not.


Abstract: Racial prejudice in the courtroom is examined through a historical sketch of racism in the legal system, a review of psychological research on White juror bias, and a study investigating White mock jurors’ judgments of a fictional trial summary. The central hypothesis is that salient racial issues at trial activate the normative racial attitudes held by White jurors. In previous eras, these racial norms encouraged overtly anti-Black prejudice. But in modern America, many Whites embrace an egalitarian value system and try to behave in an appropriately non-prejudiced manner when race is salient. Therefore, contrary to the intuition of many scholars and researchers, contemporary White jurors are more likely to demonstrate racial bias against a Black defendant in interracial trials without blatantly racial issues. Empirical data suggest that this pattern of bias is not limited to one type of crime or one type of racial issue. Practical implications and future research directions are considered.
Theory and Application:


Abstract: It is common in scientific and popular discussions to claim that unconscious racism is both prevalent and potent in modern societies. We review the theoretical models that posit different forms of unconscious racism and evaluate the empirical evidence for them. Our analysis suggests that people may sometimes lack knowledge of and control over the causes and consequences of their racial biases. However, there is little evidence to support the more provocative claim: that people possess unconscious racist attitudes. Many of the arguments to the contrary rest on strong interpretations of response patterns on implicit attitude measures. Although advances in implicit measurement can improve our understanding of racial bias, at present their use as tools for rooting out unconscious racism is limited. We describe research programs that might move these constructs to firmer scientific footing, and we urge inferential caution until such research programs are carried out.


Abstract: This article argues for the vulnerability of managerial work to unintended forms of racial and other bias. Recent insights into “implicit social cognition” are summarized, highlighting the prevalence of those mental processes that are relatively unconscious and automatic, and employed in understanding the self and others. Evidence from a response-time measure of implicit bias, the Implicit Association Test, (“IAT”; Greenwald, McGhee, and Schwartz, 1998) illustrates this phenomenon. Recent work on the predictive validity of the IAT demonstrates that social cognitive pitfalls threaten a) managers' explicit commitments to egalitarianism and meritocracy and b) managers' performance in their three primary roles of processing information, interacting with others, and making decisions (Mintzberg, 1973). Implicit bias influences managerial behavior in unexpected ways, and this influence is heightened in the messy, pressured, and distracting environments in which managers operate.

Abstract: Using a simple videogame, the effect of ethnicity on shoot/don't shoot decisions was examined. African American or White targets, holding guns or other objects, appeared in complex backgrounds. Participants were told to "shoot" armed targets and to "not shoot" unarmed targets. In Study 1, participants made the correct decision to shoot an armed target more quickly if the target was African American than if he was White, but decided to "not shoot" an unarmed target more quickly if he was White. Study 2 used a shorter time window, forcing this effect into error rates. Study 3 replicated Study 1's effects and showed that the magnitude of bias varied with perceptions of the cultural stereotype of African Americans as dangerous, and with levels of contact. Study 4 revealed equivalent levels of bias among White and African American participants in a community sample. Implications and potential underlying mechanisms are discussed.


Abstract: In this article, the authors respond at length to recent critiques of research on implicit bias, especially studies using the Implicit Association Test (IAT). Tetlock and Mitchell (2009) claim that “there is no evidence that the IAT reliably predicts class-wide discrimination on tangible outcomes in any setting,” accuse their colleagues of violating “the injunction to separate factual from value judgments,” adhering blindly to a “statist interventionist” ideology, and of conducting a witch-hunt against implicit racists, sexists, and others. These and other charges are specious. Far from making “extraordinary claims” that “require extraordinary evidence,” researchers have identified the existence and consequences of implicit bias through well-established methods based upon principles of cognitive psychology that have been developed in nearly a century’s worth of work. The authors challenge the blanket skepticism and organizational complacency advocated by Tetlock and Mitchell and summarize 10 recent studies that no manager (or managerial researcher) should ignore. These studies reveal that students, nurses, doctors, police officers, employment recruiters, and many others exhibit implicit biases with respect to race, ethnicity, nationality, gender, social status, and other distinctions. Furthermore—and contrary to the emphatic assertions of the critics—participants’ implicit associations do predict socially and organizationally significant behaviors, including employment, medical, and voting decisions made by working adults.

Abstract: This review examines the intersection of prejudice, politics, and public opinion. It focuses specifically on research that seeks to understand the sources of attitudes toward politics intended to benefit African Americans and other racial/ethnic minorities by ensuring equal treatment, providing opportunity enhancement, or striving for equal outcomes. After a review of the main patterns of white and African-American public opinion on this topic, three central theoretical interpretations of racial policy attitudes – new racism, politics and nonracial principles and values, and group conflict theories – are described and compared. The empirical evidence for each approach is assessed. Finally, directions of research that pursue a more complex view of racial policy attitudes are introduced. These include efforts to incorporate insights across theoretical domains as well as correcting an overemphasis on cognitive issues to the exclusion of affect. In addition, gaps in our understanding of “non-white” attitudes, nonprejudiced respondents, nonracial policies, and non-Americans are identified as potentially fertile ground for future research aimed at understanding the complexity of racial policy attitudes and what these can reveal about contemporary US race relations.


Author’s Summary: Building on previous research on the effects of racial priming on the opinions of White Americans, this paper engages the question of how exposure to racial cues in political messages shapes the opinions of African Americans. I argue that explanations of racial priming that focus exclusively on White Americans are insufficient to explain how racial cues influence the opinions of Black Americans, as they fail to account for the activation of in-group attitudes and mis-specify the role of explicit racial cues. In two separate laboratory experiments, I test the effects of explicitly racial, implicitly racial, and nonracial verbal cues on both Black and White Americans’ assessments of an ostensibly nonracial issue. The results point to important racial differences in the effectiveness of explicit and implicit racial verbal cues in activating racial thinking about an issue. Only frames that provide oblique references to race successfully activated racial out-group resentment for Whites. Among Blacks, explicit references to race are the most reliably elicited racial thinking by activating racial in-group identification, whereas the effect of implicit cues was moderated by the activation of negative representations of the in group. These findings not only demonstrate that racial attitude activation works differently for African Americans than for Whites but also challenge conventional wisdom that African Americans see all political issues through a racial lens.
Empirical Studies


Abstract: In 2 studies, the authors examined the role of interethnic friendship with African Americans or Latinos in predicting implicit and explicit biases against these groups. White participants completed the Implicit Association Test (IAT; Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K., 1998), several self-report bias measures, and a friendship questionnaire. Participants with close friends who were members of the target group exhibited less implicit prejudice than participants without close friends from the target group. Friendship influenced only 2 of the 7 explicit measures, a result that likely stems from social desirability bias rather than truly non-prejudiced attitudes. Results support the importance of contact, particularly interethnic friendship, in improving intergroup attitudes.


Abstract: Measures of implicit prejudice are based on associations between race-related stimuli and valenced words. Reaction time (RT) data have been characterized as showing implicit prejudice when White names or faces are associated with positive concepts and African-American names or faces with negative concepts, compared to the reverse pairings. We offer three objections to the inferential leap from the comparative RT of different associations to the attribution of implicit prejudice: (a) The data may reflect shared cultural stereotypes rather than personal animus, (b) the affective negativity attributed to participants may be due to cognitions and emotions that are not necessarily prejudiced, and (c) the patterns of judgment deemed to be indicative of prejudice pass tests deemed to be diagnostic of rational behavior.

Abstract: Many—though not all—black Americans exhibit an implicit evaluative preference for whites relative to blacks (e.g., Livingston, 2002). Are such biases meaningfully related to blacks' explicit attitudes and actual intergroup judgments? In the present study, 83 black participants who believed they would complete an intellectually challenging task with a partner rated their preferences for (fictitious) black and white potential partners. The less strongly participants implicitly preferred their ingroup, the lower their preference for a black vs. a white work partner. The magnitude of this relationship held even when controlling for explicit attitudes that were related to partner preference. Implicit biases were associated with explicit attitudes regarding black, but not white, persons and with system-justifying ideology.


Abstract: The present study investigated differences over a 10-year period in whites' self-reported racial prejudice and their bias in selection decisions involving black and white candidates for employment. We examined the hypothesis, derived from the aversive-racism framework, that although overt expressions of prejudice may decline significantly across time, subtle manifestations of bias may persist. Consistent with this hypothesis, self-reported prejudice was lower in 1998–1999 than it was in 1988–1989, and at both time periods, white// participants did not discriminate against black relative to white candidates when the candidates' qualifications were clearly strong or weak, but they did discriminate when the appropriate decision was more ambiguous. Theoretical and practical implications are considered.


Abstract: This chapter builds on the authors' concept of aversive racism as being typical of many people. Aversive racists are people who sincerely believe themselves to be unprejudiced, but who still harbor some negative feelings (often unconscious ones) toward ethnic minority groups. The authors report on a series of studies aimed at reducing people's automatic negative stereotypes about outgroups. In individual-level experiments using extensive cognitive retraining, and others creating awareness of discrepancies between one's actions and values, they demonstrated methods by which both explicit and implicit stereotypes could be reduced.
They also investigated conditions for optimal intergroup contact in which 2 groups were encouraged to recategorize their boundaries in the direction of sharing a common group identity (e.g., "we're different groups, but all on the same team.") As predicted, they found this intervention led to reduced intergroup bias and prejudice. The authors emphasize that a strong advantage of this kind of dual-identity procedure is that it does not require minority groups to forsake their own unique group identity when they adopt a broader, superordinate identity.


Abstract: The present research examined how implicit racial associations and explicit racial attitudes of Whites relate to behaviors and impressions in interracial interactions. Specifically, the authors examined how response latency and self-report measures predicted bias and perceptions of bias in verbal and nonverbal behavior exhibited by Whites while they interacted with a Black partner. As predicted, Whites' self-reported racial attitudes significantly predicted bias in their verbal behavior to Black relative to White confederates. Furthermore, these explicit attitudes predicted how much friendlier Whites felt that they behaved toward White than Black partners. In contrast, the response latency measure significantly predicted Whites' nonverbal friendliness and the extent to which the confederates and observers perceived bias in the participants' friendliness.


Abstract: Social scientists have shown that bias and stereotypes are executed and reinforced not only in moments of decision making, like hiring or promotion, but also in day-to-day interactions and social relations (or lack thereof) at work. This article argues that discrimination-reducing measures taken by employers should address the relational level in addition to the individual level of discrimination. The authors discuss existing measures for reducing employment discrimination as focusing too narrowly at the individual level of discrimination. Building on the social science literature on the influence of job segregation on interactions, they then present research supporting the idea that restructuring work away from the traditional bureaucratic model can reduce inequality by changing the context of social relations and interactions from segregated/stereotype reinforcing to integrated/stereotype challenging. They provide suggestions for several other organizational changes that may lessen discrimination at the individual level and consider the legal implications of the measures that they identify.

An implicit association test (IAT) measures differential association of 2 target concepts with an attribute. The 2 concepts appear in a 2-choice task (e.g., flower vs. insect names), and the attribute in a 2nd task (e.g., pleasant vs. unpleasant words for an evaluation attribute). When instructions oblige highly associated categories (e.g., flower + pleasant) to share a response key, performance is faster than when less associated categories (e.g., insect + pleasant) share a key. This performance difference implicitly measures differential association of the 2 concepts with the attribute. In 3 experiments, the IAT was sensitive to (a) near-universal evaluative differences (e.g., flower vs. insect), (b) expected individual differences in evaluative associations (Japanese + pleasant vs. Korean + pleasant for Japanese vs. Korean subjects), and (c) consciously disavowed evaluative differences (Black + pleasant vs. White + pleasant for self-described unprejudiced White subjects).

The present research suggests that automatic and controlled intergroup biases can be modified through diversity education. In 2 experiments, students enrolled in a prejudice and conflict seminar showed significantly reduced implicit and explicit anti-Black biases, compared with control students. The authors explored correlates of prejudice and stereotype reduction. In each experiment, seminar students' implicit and explicit change scores positively covaried with factors suggestive of affective and cognitive processes, respectively. The findings show the malleability of implicit prejudice and stereotypes and suggest that these may effectively be changed through affective processes.

Recent reports suggest that providers’ implicit attitudes about race contribute to racial and ethnic health care disparities. However, little is known about physicians’ implicit racial attitudes. This study measured implicit and explicit attitudes about race using the Race Attitude Implicit Association Test (IAT) for a large sample of test takers (N5404,277), including a sub-sample of medical doctors (MDs) (n52,535). Medical doctors, like the entire sample, showed an implicit preference for White Americans relative to Black Americans. Effects were examined among White, African American, Hispanic, and Asian MDs and by physician
gender. Strength of implicit bias exceeded self-report among all test takers except African American MDs. African American MDs, on average, did not show an implicit preference for either Blacks or Whites, and women showed less implicit bias than men. Future research should explore whether, and under what conditions, MDs’ implicit attitudes about race affect the quality of medical care.


Abstract: This study empirically examined implicit sources of bias in employment interview judgments and decisions. We examined two ethnic cues, accent and name, as sources of bias that may trigger prejudicial attitudes and decisions. As predicted, there was an interaction between the applicant name and accent that affected participants’ favorable judgments of applicant characteristics. The applicant with the ethnic name, speaking with an accent, was viewed less positively by interviewers than the ethnic named applicant without an accent and non-ethnic named applicants with and without an accent. Furthermore, modern ethnicity bias had a negative association with the favorable judgments of the applicants, which, in turn, affected hiring decisions. Implications of the results, limitations of the study, and directions for future research are discussed.