Implicit Social Cognition and Law

Kristin A. Lane,[1,2] Jerry Kang,[3]
and Mahzarin R. Banaji[2]

[1] Department of Psychology, Bard College, Annandale-on-Hudson, New York 12504;
email: lane@bard.edu

[2] Department of Psychology, Harvard University, Cambridge, Massachusetts 02138;
email: mahzarin_banaji@harvard.edu

[3] School of Law, University of California, Los Angeles, California 90095;
email: kang@law.ucla.edu

Key Words

Abstract

Experimental psychology has provided substantial evidence that the
human mind can operate in automatic, uncontrollable fashion as
well as without conscious awareness of its workings and the sources
of influence on it. With methods available to measure implicit or less
conscious aspects of social cognition, especially group-specific atti-
tudes and stereotypes, several aspects of the nature of implicit social
cognition are now regarded as well established. Such results pri-
marily include the pervasive and robust implicit favoritism for one's
own groups and socially dominant groups, the dissociation between
implicit and explicit social cognition, the ability of both to predict
behavior, the greater impact of the former on certain discriminatory
behaviors, and the sensitivity of seemingly implicit thoughts, feel-
ings, and behaviors to change in response to situational features and
experience. Legal scholarship and judicial opinions are beginning to
consider how the law can and should adapt to such findings, in par-
ticular how they call into question existing assumptions regarding
the notion of intent, and their relevance for antidiscrimination law.

[T]o be as intelligent as we can is a moral obligation—that intelligence is one of the talents for the use of which we shall be called to account—that if we haven't exhausted every opportunity to know whether what we are doing is right, it will be no excuse for us to say that we meant well.

John Erskine, *American Character and Other Essays*, 1915

In the early years of the twentieth century, John Erskine, American educator and author, worried about a lack of faith in intelligence as a virtue, and contrasted it specifically with the elevated status accorded to goodness as a virtue. To be sure, Erskine was not using the term intelligence to mean a narrowly specified mental faculty of the sort that nineteenth, and even twentieth, century psychologists called intelligence. Rather he was referring to a broad set of competencies, skills, and knowledge.

The research reviewed here is offered in an Erskinian spirit because, more now than ever, the mind sciences suggest unappealing truths about the nature of the brain and mind that originate from its bounded rationality and largely unconscious operation. Despite the commonly held belief that the opposite is true (i.e., that humans are savage rationalists and that consciousness is the default mental state), we have incontrovertible evidence that thoughts, feelings, and actions are shaped by factors residing largely outside conscious awareness, control, and intention (see Carney & Banaji 2007).

Such evidence and their implications for human nature and human experience urge that we be newly intelligent about various matters of law. Experiments from social cognition—a field concerned with the content and mechanisms of beliefs and preferences about oneself, other social beings, and social groups—are this review's mainstay, with a focus on ordinary beliefs and preferences that operate without conscious intention, awareness, or control. We present the evidence first, urging readers to heed Erskine's message when the data reveal unappealing reflections of human behavior, including our own.

THE SCIENCE

Imagine sitting at a computer. Your job appears simple: As words such as happy and angry appear sequentially on the screen, indicate whether each is good (happy is good) or bad (angry is bad) by pressing marked keys on a keyboard. But more than the words appear on the screen. In fact, each word to be judged as good or bad is preceded by a black or white face (i.e., individuals with origins in Africa or Europe) that you see but do not respond to. You merely ignore the face and respond to the words. First presented in a psychology lab over a dozen years ago, this task represents the basic method of sequential or repetition priming, designed to measure indirectly less conscious racial attitudes (Fazio et al. 1995).

The computer records the time taken to offer the easy answer that happy represents a positive or good concept and that angry represents a negative or bad concept. To the psychologists who performed this study, the data of interest were the speed to respond (with some attention to accuracy of responses) to each word. Data were sorted into four separate types: Trials in which good words like happy were preceded by (*a*) a white face, (*b*) a black face; and trials in which bad words like angry were preceded by (*c*) a white face, (*d*) a black face. From the many studies that have used this procedure, we know that speed to judge that happy is good is noticeably faster when that word is preceded by the mere flash of a white (rather than a black) face. Likewise, it is mentally easier to respond that angry is bad when it is preceded by the brief presence of a black rather than white face. This differential ease of pairing white+good and black+bad is taken as an indirect measure of the strength of automatic relative preference for the two social groups.

If the interest is in knowing a person's racial attitude, why use such an odd measure? Why not merely ask for reports of

feelings toward black and white Americans? The simple answer is that decades of research on the nature of perception, attention, memory, and decision making demonstrate that indirect measures that bypass the mind's access to conscious cognition tell us something interesting about mental states and the behaviors they spawn. Specifically, the virtue of such methods is that they tell us something different from self-reported survey-type responses. Moreover, they may potentially predict meaningful behaviors of the sort that are central to any system of law, e.g., behaviors that help and harm. For instance, the strength of black+bad and white+good associations in white subjects predicted the quality of their interaction with black experimenters (Fazio et al. 1995). Participants with stronger antiblack bias on the computerized test were less likely to be friendly toward the black experimenter than those with more positive scores. Moreover, such participants were more likely to report that blacks, compared with whites, had greater responsibility for the civil unrest, riots, and violence in Los Angeles following the 1992 acquittal of police officers in the case involving Rodney King.

Interestingly, the same participants' self-reported attitudes toward the groups on a traditional survey were uncorrelated with their friendliness; in other words, these traditionally measured expressions of attitude were not as predictive of behavior. This experiment set the stage for the subsequent explosion of work on implicit social cognition—the main results of Fazio and colleagues (1995) would be replicated many times with differing measures of implicit attitudes and stereotypes. This body of research captured the essence of a new generation of discoveries about automatic, nonconscious, or implicit preferences and beliefs, primarily that they (*a*) are both pervasive (large numbers of individuals show evidence of them) and large, statistically speaking; (*b*) diverge from the consciously reported preferences and beliefs of the same individual; (*c*) appear to predict behavior, even consequential behavior such as doctors' treatment

of patients (Green et al. 2007); and (*d*) despite their seemingly uncontrollable nature, are malleable.

Two concepts, attitude (or preference) and stereotype (or belief), are central to the study of implicit social cognition and its application to the law. "*Implicit attitudes* are introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (Greenwald & Banaji 1995, p. 8). For example, an object is likely to be preferred as a result of prior exposure to it, even without awareness that such exposure causes the increase in preference (Zajonc 1980). Likewise, "*Implicit stereotypes* are the introspectively unidentified (or inaccurately identified) traces of past experience that mediate attributions of qualities to members of a social category" (Greenwald & Banaji 1995, p. 15). An implicit stereotype would be the (mistaken) identification that Dave Sebastian is famous, but Diane Sebastian is not (Banaji & Greenwald 1995), based on the belief (a correct belief, in this case) that men are more likely to be famous than women. Such preferences (attitudes) as well as the ascription of specific qualities (stereotypes) are jointly referred to here as implicit social cognitions (ISCs), and the biases observed in studying them are referred to as implicit biases.

As an example of how the science of implicit social cognition provides an alternate lens for viewing human behavior, consider two police shootings in the outer boroughs of New York City. Both Amadou Diallo (in 1999) and Michael Bell (in 2006) were unarmed black men, mistakenly shot and killed by police officers. Of central relevance is the probability of the occurrence of such a response had the victims been white rather than black. In the earlier case, officers reported mistaking Diallo's reach into his pocket as an attempt to get a gun. During the opening statements in the criminal trial of the four police officers—who were all acquitted—the prosecution argued that "when they got out of the car in

Access provided by Bard College on 03/06/17. For personal use only.

front of Amadou Diallo's home in the early morning of February 4, they made the conscious decision to shoot him."

Data and theory from implicit social cognition provide a different interpretation of these incidents, namely that the officers acted without conscious racial animus but were still influenced by the victims' race. No satisfactory understanding of the role of implicit bias in the police officers' behavior can be obtained by analyzing single lethal events, for we cannot know whether the same officers in an identical universe but for the victim's race would have acted differently. The laboratory offers a way to replicate the essential conditions of such situations and to examine patterns of behavior that arise in systematic ways. Several laboratories created controlled environments analogous to the police officers' situation. In one such study, black and white men appeared one at a time on a computer screen, holding either a gun or a harmless object (e.g., a soda can) (Correll et al. 2002). If the target held a gun, participants were instructed to press one key to shoot; if it was a harmless object, they were told to press another key for don't shoot.

The data revealed systematic racial bias in shooting, with faster and more accurate responses to unarmed white targets and armed black targets compared with armed white targets and unarmed black targets. Neither participants' endorsement of racial stereotypes nor their reports of feelings toward blacks predicted shooter bias. Knowledge of cultural stereotypes, however, did predict shooter bias: Those with greater awareness of the stereotype of African Americans as hostile were more likely to mistakenly shoot unarmed black suspects and not shoot armed white suspects (see Greenwald et al. 2003b, Payne 2001, Plant & Peruche 2005 for similar results). In light of these data, the New York City shootings can be explained without reliance on explicit bias. Simple exposure to the stereotype that blacks are hostile, even without endorsement of that stereotype, may be sufficient to create bias that alters split-second

decisions and does so without conscious awareness.

Among the criticisms of such studies is that they reflect a reality far from that of the professionals who rely on their gut to make the right decisions in their real jobs—college students, after all, don't spend their hours in target practice. To test a different group of professionals, those who take an oath to serve the sick whether rich or poor, we studied the behavior of physicians making assessments of patients. Green and colleagues (2007) measured implicit bias among emergency room physicians. Physicians recommended treatment based on vignettes that depicted patients with myocardial infarction who differed solely on race. Although doctors' reports of racial attitudes and beliefs did not predict their proposed treatment, their ISCs did: Physicians with stronger implicit antiblack attitudes and stereotypes were less likely to prescribe thrombolysis (a blood-thinning procedure) for African Americans compared with white Americans with identical medical profiles. Explicit race attitudes, held with complete sincerity and showing no race bias, did not predict medical recommendations.

Consciously held attitudes and stereotypes are also important predictors of behavior. They are simply not the only ones to contend with as we understand human behavior and its vicissitudes. We can study ISCs with physiological techniques that measure cardiovascular responses (e.g., Blascovich et al. 2001) or micro facial movements (e.g., Vanman et al. 2004), neuroimaging techniques that measure brain activation (e.g., Mitchell et al. 2006, Phelps et al. 2000), and those based on response time, such as the priming task described earlier (Perdue et al. 1990), to measure automatic attitudes and beliefs. This last class has thus far been the most reliable and widely used for measuring individual differences in ISCs and, consequently, represents the bulk of the work reviewed here.

The logic underlying such measures is relatively straightforward. First, a very simple definition of an attitude is assumed:

"Essentially, then, an attitude can be viewed as an association between a given object and a given evaluative category" (Fazio et al. 1982, p. 341). Second, response speed to varying stimuli measures the association's strength. In the priming task described earlier, those with antiblack feelings were presumed to have a relatively stronger association between the social category black and the evaluative category bad (compared with the categories white and good). The difference in this response time is presumed to reflect a person's attitude toward blacks (see Wittenbrink 2007 for further information on the priming task).

Other measures involve similar logic but differ in implementation. The Implicit Association Test (IAT; Greenwald et al. 1998), for example, requires participants to rapidly classify individual stimuli into one of four distinct categories using only two responses. As with priming, the assumption is that responses will be facilitated—and thus faster and more accurate—when categories that are closely associated are paired than when they are not paired (see Lane et al. 2007, Nosek et al. 2006 for further information on the IAT). For example, a person with a negative implicit attitude toward blacks would be expected to go more quickly when black and bad share one key and white and good the other than when the pairings of good and bad are switched (readers may try an IAT at **http://www.projectimplicit.com**). These measures—priming and the IAT—are the most commonly used tasks to measure ISCs and are emphasized in our review, although newer tasks, such as the Go/No-Go Association Task (Nosek & Banaji 2001), the Evaluative Movement Assessment (Brendl et al. 2005), the extrinsic affective Simon task (De Houwer 2003), and the affect misattribution procedure (Payne et al. 2005), have been used, and likely more will be developed.

Other tasks measure ISCs without measuring response speed or physiological response. Completion of word fragments may reveal mentally active stereotypes [e.g., _OOR might be "POOR" when social class is on one's mind, but "DOOR" if it is not (e.g., Sinclair & Kunda, 1999)]. Additionally, the extent to which behaviors completed by black actors are described in abstract terms (von Hippel et al. 1997) or to which stereotype-inconsistent behaviors are explained rather than just described (Sekaquaptewa et al. 2003) is used as a measure of implicit stereotypes.

This report draws four key conclusions about the current state of knowledge of ISCs: (*a*) Measures that assess ISCs provide distinct assessments from self-reported or explicit versions; (*b*) preferences for ingroups as well as socially valued groups are widespread; (*c*) ISCs relate systematically to behavior; and (*d*) ISCs are flexible and respond to experience and environmental cues. The practical and ethical implications of such findings have already been explored in preliminary ways in various domains [Banaji et al. 2003, Banaji & Bhaskar 2000, Bazerman et al. 2005; see also the 2006 *California Law Review* symposium on behavioral realism (Bayern 2006)]. In this review, we focus on how these scientific findings have influenced legal scholarship and judicial opinions thus far.

Implicit Social Cognition is Distinct from Explicit Social Cognition

ISCs often reveal different levels of intensity and patterns of attitudes and stereotypes than those provided by explicit, self-report measures. Evidence suggests that implicit and explicit measures provide distinct, although sometimes related, assessments of biases.

For example, although people tend to report only slight preference for white Americans over black Americans, implicit measures show a quite different tendency: On several implicit measures, strong and consistent preferences emerge for white Americans relative to black Americans (e.g., Nosek et al. 2002, 2007). Such data, in conjunction with small correlations between explicit and implicit measures of the same attitude or stereotype,

support the idea that the two systems of implicit and explicit social cognition exist as separate mental spheres with communication channels that are present but don't always work.

Relationships between implicit and explicit measures vary across study, target group, and participant characteristics (Bosson et al. 2000, de Jong et al. 2003, Egloff & Schmukle 2002, Greenwald et al. 1998, Karpinski & Hilton 2001, Ottaway et al. 2001, Rudman & Kilianski 2000). A meta-analysis—a comprehensive quantitative analysis of experiments on a particular topic allowing more general conclusions than any single study—of 126 correlations between implicit (assessed with the IAT) and explicit attitudes revealed considerable variability in the strength of the relationship between implicit and explicit cognitions, although on average they were related to one another, mean population $r = 0.24$ (Hofmann et al. 2005).

Statistical analyses of responses to both measures further support the idea that implicit and explicit measures tap separate processes. A construct represents an abstract concept that, although measurable, cannot be directly observed, such as self-esteem, IQ, or in this case implicit and explicit social cognitions. Do implicit and explicit measures provide assessments of a single or multiple constructs? A statistical procedure, confirmatory factor analysis (CFA), tests whether measures tap the same underlying construct (in which case they would, in CFA's language, load onto a single factor) or multiple constructs (in which case they would load onto multiple factors). Consistent with the notion that implicit and explicit attitudes are distinct (Greenwald & Banaji 1995, Wilson et al. 2000), implicit and explicit attitudes are best fit as separate factors, even when they are correlated. For example, implicit self-esteem and gender identity each loaded onto a separate factor than their explicit counterparts (Greenwald & Farnham 2000). Implicit racial attitudes (measured by the IAT and priming) also represented a separate construct from explicit racial attitudes

(Cunningham et al. 2001). This pattern is more general: Implicit and explicit attitudes were best represented by two factors for 56 out of 57 different attitude objects (Nosek 2005).

Evidence from the neural basis of social cognition converges with that from behavioral tests. The focus has been on subcortical and cortical structures, with the loose assumption that implicit evaluations that are less controllable should map onto activations in the former and processes that rely on deliberative thought should engage the latter. The amygdala is a subcortical brain structure known to be reliably engaged in processing emotional (especially fear-relevant) and novel stimuli (Phelps 2006). When faces were presented so rapidly as to be under the limen or subliminal threshold (30 milliseconds, or 3/100th second), black faces elicited greater amygdala activation than did white faces. When faces were exposed long enough to be visible to participants (525 milliseconds) no such difference in amygdala activation was observed, and instead greater activation was seen in regions assumed to be associated with regulating thoughts and exerting control (the right ventrolateral prefrontal cortex, right dorsolateral prefrontal cortex, and anterior cingulate). Moreover, amygdala activation correlated with the IAT measure of racial attitudes when the faces were presented subliminally, but not so when presented supraliminally, indicating that the IAT reflects more automatic rather than controlled reactions to social groups (Cunningham et al. 2004a).

Most experts (these authors included) do not believe that measures of implicit social cognition reflect the "true" attitude any more than do measures of explicit social cognition such as questionnaire responses (Fazio & Olson 2003, Lane et al. 2007). Implicit and explicit measures appear to tap separate constructs that operate differently: They both predict behavior (which one predicts better appears to depend on the person and situation). Privileging one over the other would be scientifically misguided.

Implicit Social Cognitions are Robust and Pervasive

One of the most consistent findings from the large literature on intergroup relations is the fact of ingroup favoritism (Tajfel & Turner 1986). This tendency is so strong that people report liking ingroups even when they are randomly assigned to them (Cadinu & Rothbart 1996, Gaertner et al. 1989, Tajfel et al. 1971). This pattern also emerges on implicit measures: After random assignment to the group "Quan" or "Xanthie," participants demonstrated implicit preference for their assigned group (Ashburn-Nardo et al. 2001). It is mirrored on measures of implicit preference for known social groups. For example, both Japanese Americans and Korean Americans preferred their own ethnic group relative to the other (Greenwald et al. 1998), as did East and West Germans (Kühnen et al. 2001).

Positive implicit attitudes toward and stereotypes about members of socially privileged groups are also pervasive. At the website mentioned above (**http://www.projectimplicit.com**), visitors try one (or more) IATs and receive feedback about the magnitude and direction of their implicit attitude or stereotype. With over 5 million tests completed, this is the largest repository of data available to look at variability and frequency of ISCs (see Nosek et al. 2007 for a review). Of course, visits to the website are optional, and thus the data do not reflect a representative sample of the population. Even with this caveat, the data provide a unique look at ISCs and allow comparisons across self-reported demographic characteristics (e.g., age, gender, race/ethnicity, region, political orientation).

Table 1 depicts results from 17 IATs available at the website (Nosek et al. 2007). Two features of the data are readily apparent. First, implicit preferences in either direction, away from the neutral position of no bias, were typical. On average, participants preferred socially privileged groups (young over old, white over black, light-skinned people over dark-skinned people, other people over Arab Muslim, abled people over disabled people, thin people over obese people, and straight people over gay people). Implicit stereotypes were also consistent with widespread cultural beliefs. For example, most participants (72%) associated the concepts male with science and female with humanities. Similarly, participants found it easier to categorize white, rather than Asian or Native American, faces with American, reflecting an implicit stereotype that "American equals white" (Devos & Banaji 2005; T. Devos, B.A. Nosek & M.R. Banaji, unpublished manuscript). Additionally, participants exhibited stereotypes reflecting associations between blacks and weapons (compared with whites and harmless objects), and between the social group male and career (compared with female and family). Explicit reports of ingroup preference and stereotypes were smaller in magnitude in white Americans than those observed on implicit measures.

Second, the standard deviations shown in **Table 1** reveal that despite the prevalence of implicit bias, there was substantial variability in the extent to which people showed such bias—some people were much higher than the average (reflecting strong bias), but others were lower than the average (reflecting minimal bias or even biases in the opposite direction than the majority of test takers). For example, although 68% of test takers implicitly preferred white compared with black, a nontrivial group—14%—showed the opposite, and preferred black over white, and 18% showed no preference. Such variability suggests that additional factors may moderate an individual's level of implicit bias. Group membership attenuated implicit bias. Although people of European, Asian, and Hispanic descent implicitly preferred white over black, black participants did not, on average, show ingroup preference; equal numbers of black participants preferred the outgroup white as preferred the ingroup black. The pattern of stronger ingroup preference among members of socially privileged groups is predicted

Table 1 Attitudes and stereotypes for 17 topics among visitors to a public website: overall sample (implicit and explicit attitudes and stereotypes), by gender (implicit), and by ethnicity (implicit)[a]

Higher numbers represent stronger associations between	Overall Implicit[b]			Overall Explicit			By Gender Implicit		By Ethnicity Implicit						
	M	SD	d	M	SD	d	Women	Men	American Indian	Asian	Black	Hispanic	White	Multi-Racial	Other
Attitude															
Age — Young+Good/Old+Bad	.49	.39	1.23	.39	.78	.51	1.17	1.40	1.22	1.22	1.40	1.23	1.25	1.17	1.25
Race — White+Good/Black+Bad	.37	.43	.86	.26	.73	.36	.79	.93	.79	.88	−.05	.79	1.00	.56	.70
Skin-tone — Light Skin+Good/Dark Skin+Bad	.30	.41	.73	.17	.67	.25	.71	.80	.59	.76	.22	.68	.88	.49	.66
Child-race — White+Good/Black+Bad	.33	.45	.73	.19	1.30	.15	.69	.80	.53	.76	−.16	.67	.87	.40	.64
Arab-Muslim — Other Peoples+Good/Arab Muslims+Bad	.14	.42	.33	.45	.77	.58	.24	.48	.29	.17	.19	.36	.40	.19	.00
Religion — Other Religions + Good/Jewish+Bad	−.15	.44	−.34	.14	1.05	.13	−.36	−.34	−.20	−.14	−.18	−.20	−.39	−.30	−.36
Disability — Abled+Good/Disabled+Bad	.45	.43	1.05	.38	.67	.57	.98	1.28	.98	1.09	1.14	1.02	1.07	.98	.93
Sexuality — Straight People+Good/Gay People+Bad	.35	.47	.74	.49	.91	.54	.68	.94	.83	.91	1.11	.87	.74	.66	.70
Weight — Thin+Good/Obese+Bad	.35	.42	.83	.64	.73	.88	.81	.93	.81	.76	.74	.81	.88	.76	.79
Presidential — Bush+Good/Other President+Bad[c]	−.07	.45	−.15	−.94	1.28	−.73	−.20	−.13	−.18	−.24	−.33	−.20	−.16	−.27	−.24
Election 2004 — George Bush+Good/John Kerry+Bad	−.14	.51	−.27	−.69	1.64	−.42	−.37	−.20	−.25	−.37	−.61	−.31	−.27	−.37	−.39
Election 2000 — George Bush+Good/Al Gore+Bad	−.09	.56	−.16	−.32	1.60	−.20	−.20	−.07	−.11	−.32	−.70	−.23	−.11	−.25	−.25
Stereotypes															
Race-Weapons — White+Harmless Objects/Black+Weapons	.37	.37	1.00	.34	1.10	.31	.97	1.03	.84	.92	.59	.92	1.05	.86	.89
American-Native — White American/+American/Native American/Foreign	.23	.50	.46	−.76	1.79	−.42	.46	.48	−.24	.48	.44	.46	.56	.16	.30
American-Asian — White+American/Asian+Foreign	.26	.41	.62	.57	1.27	.45	.61	.68	.59	.29	.83	.68	.85	.51	.61
Gender-Science — Male+Science/Female+Liberal Arts	.37	.40	.93	.52	.66	.79	.98	.93	.90	.88	.85	.85	.98	.88	.93
Gender-Career — Male+Career/Female+Family	.39	.36	1.10	.54	.60	.89	1.19	.94	1.06	.97	1.19	1.06	1.14	1.06	1.03

[a]Adapted from Tables 1, 2, 3, and 4 of Nosek et al. (2007).

[b]IAT scores are mean D scores (Greenwald et al. 2003a). d = Cohen's d effect size, which is a standardized measure of the magnitude of an effect; the following guidelines indicate the size of the effect: small d = 0.2, medium d = 0.5, large d = 0.8 (Cohen, 1988).

[c]The comparison category for George Bush varied (Clinton, Reagan, Nixon, Kennedy, FDR, Lincoln, Jefferson, or a collection of recent presidents).

by system justification theory (Jost & Banaji 1994), which suggests that the tendency to maintain the status quo will lead to reduced ingroup preference among members of disadvantaged groups (see Jost et al. 2004 for a review of system-justifying tendencies on implicit and explicit measures).

Although the tendency toward ingroup liking appears strong, lab studies also demonstrate that a group's status moderates the magnitude of implicit ingroup preference. Predominant cultural evaluations attenuate this tendency toward ingroup preference (see Jost et al. 2004 for a review). As in the web data summarized in **Table 1**, although most white Americans showed strong preference for white over black on the IAT, black Americans, on average, did not prefer either group (Ashburn-Nardo et al. 2003, Livingston 2002, Nosek et al. 2007). Similarly, weight and socioeconomic status (Rudman et al. 2002), as well as university (Jost et al. 2002) and college dorm status (Lane et al. 2005) moderated the strength of ingroup liking, such that people from higher-status groups were more likely to exhibit implicit ingroup preference.

In other words, the elderly, the poor, and those associated with less prestigious institutions all showed weaker preference for their own group on implicit measures. Although the tendency toward ingroup preference is robust, a group's status in the larger social hierarchy is a decisive influence on the implicit biases that are observed. Members of privileged groups overwhelmingly show ingroup preference (70% and up is not a rare finding), whereas this tendency is subdued among members of socially derogated groups who internalize the broader cultural evaluation of their group. The elderly, for example, show no ingroup-favoring implicit attitudes; gays and lesbians show weaker ingroup preference than do heterosexual participants (Nosek et al. 2007); and blacks in the United States and in South Africa show substantially weaker ingroup preference than do whites in those nations (Shutts et al. 2007).

A claim has been made that ISCs, in particular preferences for social groups, reflect not a person's own attitude but rather knowledge that the person has acquired about the attitude present in the larger culture (Karpinski & Hilton 2001). Of many possible responses to this claim, we offer three observations that suggest that ISCs do reflect the state of an individual's mind. First, implicit preferences emerge even without cultural knowledge about the group: Simply being told you are a Quan is sufficient to generate implicit preference for Quans (Ashburn-Nardo et al. 2001). If implicit biases solely reflect knowledge of cultural hierarchies and stereotypes that exist separately from individual social cognition, preferences favoring these nominal groups should not appear.

Second, as reviewed more extensively below, the magnitude of implicit bias predicts behavior (Poehlman et al. 2007); if implicit biases reflect disjointed cultural knowledge, they should not be linked to a specific individual's discriminatory behavior. Finally, we note that it is often explicit, rather than implicit, attitudes that systematically relate to reports of cultural knowledge (Nosek & Hansen 2007). The culture and the person are intricately intertwined, and it would be a mistake to assume that signals of ingroup-favoring attitudes, because their content may have its origins in culturally shared knowledge, are hence not a part of the individual (Banaji 2001). Just as it would make little sense to ask what makes a rectangle a rectangle, its length or its width, so too would disambiguating person from culture in any simplistic way be nonsensical.

Implicit Social Cognitions Predict Behavior

The nature of ISCs should be of interest to the law to the extent that they predict behavior. Before such an exercise is initiated, it is important to know if implicit measures differentiate groups known to differ a priori. Indeed, in several domains, the IAT detects expected group differences. Several known-groups

validations exist from the clinical and social domains. Although both criminals high and low in psychopathy [characterized by "grandiosity, callousness, manipulation, lack of empathy, and lack of guilt or remorse" (Snowden et al. 2004, p. 621)] implicitly preferred the concept "peaceful" compared with "violent," this preference was muted among murderers diagnosed as psychopathic (Gray et al. 2003, Snowden et al. 2004); that is, they exhibited lower implicit dislike for violence than did nonpsychopathic murderers. Likewise, clinically identified pedophile criminals implicitly associated images of children with sex, whereas nonpedophile criminals associated adult images with sex; 78% of pedophiles in this study were correctly classified by the task (Gray et al. 2005). An implicit measure clearly distinguished between subjects who had phobias associated with either snakes or spiders (Teachman & Woody 2003) and between subjects who were smokers and nonsmokers (Swanson et al. 2001). The IAT also detected stronger self+injury associations in self-injurers, as well as suicide ideation in those who are at risk (Nock & Banaji 2007a, Nock & Banaji 2007b). It also efficiently demarcates social groups. Men and women are distinguishable with near-100% accuracy on a gender identity measure. Other tests also detect group differences such as Japanese versus Korean, black versus white, and gay versus straight, among others (Nosek et al. 2007).

Importantly, implicit bias also predicts individual discriminatory behaviors. In addition to predicting medical interns' treatment decisions (Green et al. 2007), people with greater implicit negativity toward blacks were less likely to anticipate befriending an African American and tended to sit further away from an African American partner. Those who more strongly associated black with the concept "physical" (compared with the concept "mental") were also more likely to predict that an African American partner would perform poorly on a joint academic task and rate an individual African American partner in more stereotypic fashion (Amodio & Devine 2006).

A priming measure assessing attitudes toward the overweight also predicted how far participants placed their chair from an overweight woman (Bessenoff & Sherman 2000). Strikingly, white students' implicit attitudes toward blacks, measured at the beginning of the semester, predicted the longevity of their relationship with a randomly assigned black roommate. Those who showed the most implicit negativity toward blacks at the school year's start were more likely not to be living with their black roommate at the academic year's end (Towles-Schwen & Fazio 2003). These results sit in line with laboratory findings that people with more negative implicit racial attitudes were viewed as less friendly by confederates or observers unaware of their racial attitudes (Dovidio et al. 2002, Fazio et al. 1995, McConnell & Leibold 2001).

Nonverbal behaviors such as facial expressions, eye contact, and body posture have been shown to leak implicit attitudes. Those who possess stronger negative attitudes toward a stigmatized group tend to exhibit more negative behaviors (e.g., blinking) and less positive behaviors (e.g., smiling) when interacting with a member of that group (Lemm 2006, McConnell & Leibold 2001). Further, ISCs relate to judgments of identical actions or objective states, depending on the target's group membership. People with higher implicit bias judged ambiguous actions by a black (Rudman & Lee 2002) or Turkish (Gawronski et al. 2003) target more negatively. Even the same facial expression appeared different depending on implicit bias. More negative implicit racial attitudes were related to a lowered threshold for detecting hostility on black, but not white, faces (Hugenberg & Bodenhausen 2003).

Other lines of research demonstrate the relationship between ISCs and lower-level behaviors that are not immediately observable, such as cognitive processing or neurological responses. Implicit racial attitudes assessed by the IAT were linked to greater amygdala activation when viewing unfamiliar black (versus white) faces (Cunningham et al. 2003, Phelps

et al. 2000). Additionally, white participants with stronger antiblack bias performed poorly on a measure of cognitive control after interacting with a black partner, suggesting that those with stronger bias used more cognitive resources during the interaction (Richeson & Shelton 2003). In another study, participants viewed unfamiliar black faces while activation in key brain regions was assessed. Those with stronger antiblack bias exhibited more activation in brain regions associated with cognitive control, suggesting they were engaging in attempts to control automatic reactions to the faces. Activation in one of these regions—the dorsolateral medial prefrontal cortex—mediated, or accounted for, the relationship between implicit bias and amount of cognitive interference on a task that followed an interaction with a black individual (Richeson & Shelton 2003). These data suggest that the depletion of cognitive resources while interacting with a black partner was due to an attempt to exercise control over one's biases.

The most systematic exploration of the connection between implicit bias and behavior is a meta-analysis of studies that included the IAT and participant behaviors (Poehlman et al. 2007). Across studies, ISCs predicted a range of criterion variables, including nonverbal behavior, social judgments, physiological responses, and social action. Both implicit (average $r = 0.27$) and explicit (average $r = 0.34$) measures were significantly related to behavior across topics. In the domain of stereotyping and prejudice, ISCs better predicted criterion behavior (average $r = 0.22$) than explicit cognitions (average $r = 0.12$).

Moderators of the bias-behavior relationship. Implicit bias does not inevitably lead to behavior. With the relationship between bias and behavior fairly well established, research is turning to when and how implicit bias is likely to be linked to action. In particular, motivation may moderate the link between ISCs and behavior. The MODE (Motivation and Opportunity as DEterminants of behavior) model of the relationship between attitude

and behavior (Fazio & Towles-Schwen 1999) suggests that both motivation to be egalitarian and the opportunity to control one's behavior affect whether implicit bias is manifested behaviorally. Consider two people with identical negative implicit outgroup attitudes. If person A is motivated to be nonbiased and able to control the influence of his bias, then the MODE model suggests that the ISC-behavior link will be disrupted. Alternatively, if person B lacks such motivation, her ISCs are more likely to influence behavior.

Data support this idea. Among participants low in motivation to control bias, implicit bias (measured via priming) predicted trait ratings of black targets, relative to whites. Those highly motivated to control bias showed the reverse pattern, indicating they may have been overcorrecting for their implicit bias (Olson & Fazio 2004). Similarly, for white participants low in motivation to control prejudicial responses, implicit bias predicted anticipated comfort level during an unscripted interaction with a black partner. However, it was not predictive for those strongly motivated to control prejudice (Towles-Schwen & Fazio 2003). Implicit bias toward gay people predicted nonverbal behavior during an interaction with a gay partner only for participants low in motivation and in the tendency to control their behavior (Dasgupta & Rivera 2006).

The path from implicit bias to negative behavior does not appear immutable. At least under certain conditions, when people are motivated to behave in a nonprejudiced manner, they may be able to override the effects of implicit bias on behavior. People are most likely to be able to perform this correction when they can recognize the potential for social group membership to influence their behavior and to control the behavior itself.

Implicit Social Cognitions are Malleable

Despite their prevalence and magnitude, ISCs are not impervious to change. Like their explicit counterparts, the immediate situation's

demands and a person's personality affect the strength of ISCs. Situational features affect implicit biases, either by altering some feature of the target social group or by changing the participant's momentary goals or motivations (see Blair 2002 for a review). Implicit bias not only changes as a function of the social situation but also varies with personal characteristics.

Implicit biases are sensitive to features of the local situation. Exposure to counter-stereotypical outgroup members often reduces implicit bias. For example, implicit race bias was reduced both immediately and 24 hours following exposure to well-liked African Americans and disliked white Americans (Dasgupta & Greenwald 2001), whereas imagining a female leader reduced implicit gender stereotyping (Blair et al. 2001). Additionally, when the group "black" was represented by well-liked African Americans such as Michael Jordan, implicit attitudes toward blacks became more positive (Mitchell et al. 2003).

Benefits of exposure to (or thoughts about) counterstereotypical individuals extend beyond the lab: Dasgupta & Asgari (2004) measured implicit gender stereotypes (male+leader) among female students at a coeducational and an all-women's college. Although women at both schools held similar implicit beliefs at the start of college, after one year women at the single-sex college did not display any implicit bias, whereas women at the coeducational school had, on average, stronger implicit stereotypes of male+leader. Greater exposure to female teachers at the single-sex school was responsible for the decrease in bias. Presumably, observing and interacting with female professors strengthened the association of female+leader.

Conscious exertion to be unbiased may—at least temporarily—reduce implicit bias. In one study, participants responded "no" when they saw a social group (skinheads) paired with a stereotypic trait, and "yes" when they saw the group paired with a counterstereotypic trait. After 480 trials of saying "no" to stereo-types or "yes" to counterstereotypes, implicit stereotyping was eliminated (Kawakami et al. 2000). Similarly, repeated pairings of black faces with positive words during an ostensibly unrelated exercise resulted in more egalitarian implicit racial attitudes, even though participants were unaware of any systematic pairing between positive words and black faces. This reduction persisted for two days following exposure to the black-positive pairing (Olson & Fazio 2006). Racial shooting bias in a police simulation was decreased after repeated exposure to pairs of stimuli in which ethnicity was unrelated to criminality (Plant et al. 2005).

Besides these interventions, other techniques have capitalized on the extent to which implicit bias depends on salient social categories. Michael Jordan, like mere mortals, belongs to multiple social categories: gender (male), occupation (athlete), and race (African American). Classifying prominent black athletes and white politicians according to their occupation reduced antiblack implicit bias compared with classification by race (Mitchell et al. 2003; see Barden et al. 2004 for a similar result). In a striking example of how multiple identities can shape performance, a subtle reminder of Asian women's gender or ethnicity lowered or raised their performance, respectively, on a math task, compared with a control condition (Shih et al. 1999). Even abilities expected to be stable and impervious to such minor suggestions are affected by factors not easily identifiable.

Goals vary by situation, and immediate goals help to determine ISCs. Implicit stereotypes were reduced after receiving positive feedback from a black manager (presumably increasing motivation to value him) but increased after receiving negative feedback (presumably increasing motivation to derogate him) (Sinclair & Kunda 1999). Desire to affiliate with others lowers bias: Implicit racial bias was lower after interacting with a black superior than with a black subordinate (Richeson & Ambady 2001) or after interacting with an experimenter who wore a shirt that said "Eracism" (implying egalitarian beliefs) when

the experimenter was well liked (Sinclair et al. 2005).

Motivation, of course, is not solely situational: Some people are more dispositionally motivated to be nonprejudiced, a tendency that moderates ISCs. People motivated to be nonprejudiced for personal (or internal) reasons, but not social (or external) reasons showed reduced implicit racial bias on a physiological measure (Amodio et al. 2003) and a reaction-time task (Devine et al. 2002; see Vanman et al. 2004 for a case in which motivation was related to a reaction time, but not to a physiological, measure of bias). Implicit bias is also related to more general cognitive styles, such that people with highly rigid thinking styles or strongly right-wing ideologies exhibit stronger implicit bias (Cunningham et al. 2004b).

Personal and situational factors do not exist in a vacuum, and the individual traits that a person brings to a situation often interact with the immediate context. Social dominance orientation (SDO), an attitudinal system referring to whether people prefer hierarchically organized groups, was related to implicit ingroup preference only when the students' university was threatened. Without social threat, SDO was unrelated to preference for a generic ingroup. Under a condition of group threat, participants who preferred hierarchical arrangements of social groups (high-SDO) showed much larger implicit ingroup preference than did low-SDO students (Pratto & Shih 2000). Peruche & Plant (2006) found that training participants not to associate race with athleticism (by repeated exposure to pairings of black and white faces with athletic or nonathletic objects, in which race and athletic features were independent) was particularly successful in reducing this race-athlete stereotype among participants highly motivated for personal reasons to be nonbiased.

THE LAW

Implicit biases appear to be widespread, to be discrepant from self-reported bias, to in-

fluence behavior, and to be sensitive to intervention. What are the legal implications? Broadly framed, this question is not novel. Legal scholars have long analyzed the significance of various states of mind, including ones that are not entirely purposeful; the topic of unconscious prejudice has been discussed for decades (Lawrence 1987). Recent findings in implicit social cognition sharpen the question because it is now supported by replicable evidence from numerous laboratories rather than by anecdote, hunch, or psychoanalytic theory.

Indeed, assuming that the scientific research continues along its current trajectory, implicit social cognition has the potential to influence the understanding of intent in all bodies of law. For instance, constitutional and statutory law governing civil rights and the equal treatment of individuals is clearly subject to revision because implicit social cognition destabilizes conventional understandings of disparate treatment, disparate impact, hostile environments, and color or gender consciousness. Observed disparities between social groups will likely be examined for connection to implicit bias as an aggravating factor. In criminal law, the data are relevant in debating policies and laws surrounding racial profiling, self-defense, community policing, jury selection, and penalty setting. Finally, the various media that transmit and sustain our culture will increasingly become a source for investigation of implicit bias because they are the obvious source of what we see and hear. Already, such matters have begun to be examined in law reviews and cases.

Legal Literature

The modern history of legal scholarly engagement with implicit social cognition began in the mid-1990s. In 1995 and 1998, Krieger wrote the seminal articles applying cognitive psychology to questions of, first, employment discrimination and, next, affirmative action. She argued that employment discrimination may be caused not by racist villains who

relish inflicting harm on minorities; rather, it may be a natural byproduct of banal cognitive sorting (Krieger 1995). Affirmative action may exacerbate negative stereotypes of minority beneficiaries. However, a strategy of color-blindness is cognitively naive (Krieger 1998). Written largely before the empirical focus on individual differences in ISCs, these articles relied on the antecedent psychological literature on bounded rationality and emphasized the role of schematic thinking. Nonetheless, they created the foundation for subsequent legal analyses. Other significant contributions drawing on similar science include Pollard's (1999) call for an evidentiary privilege for employers who test for unconscious bias and Armour's (1995) call to break the prejudice habit.

By 2002, legal scholars began focused engagement with implicit social cognition science, including the IAT, which by then had gained scientific prominence. That year, Blasi (2002) described various psychological theories with substantial attention to automatic categorization, motivated cognition, and implicit bias, and analyzed why advocacy based on folk theories of prejudice would fail. In particular, he rejected traditional advocacy strategies that understood stereotypes as either empirical mistakes to be corrected or moral failures to be redeemed.

In 2003, Saujani provided extensive discussion of the IAT and suggested that it could play a useful role in adjudicating equal protection claims. In *Washington v. Davis* (1976), the Supreme Court clarified that an equal protection violation required a finding of discriminatory intent, not mere disparate impact. Because such intent may have to be established through circumstantial evidence, the Court identified various relevant factors in a subsequent case, *Village of Arlington Heights v. Metropolitan Housing Development Authority* (1977). Saujani (2003) suggested that legislators should take the IAT and have those results admitted as relevant evidence in this search for discriminatory intent. Although such a recommendation far exceeds what the lead-

ing scientists of the IAT, including its creators, view as responsible use, Saujani raised thoughtful, provocative questions.

In 2005, Kang provided arguably the first systematic synthesis in the law reviews of the implicit bias research, in support of a model of racial mechanics. In addition to providing this synthesis, Kang (2005) questioned the Federal Communications Commission's strong preference for local news as the way that broadcasters should satisfy the Communication Act's public interest standard. Given the violent crime stories disproportionately featured on local news, he queried whether news programs functioned as Trojan horse viruses that increased implicit bias against minorities. He also suggested a broad legal research agenda and was the first to encourage a behavioral realist incorporation of the science of implicit social cognition.

In 2006, the *California Law Review* published a special symposium volume on behavioral realism (Bayern 2006). In simplified terms, behavioral realism involves a three-step process. First, identify advances in the mind and behavioral sciences that provide a more accurate model of human cognition and behavior. Second, compare that new model with the latent theories of human behavior and decision making embedded within the law. These latent theories typically reflect common sense based on naive psychological theories. Third, when the new model and the latent theories are discrepant, ask lawmakers and legal institutions to account for this disparity. An accounting requires either altering the law to comport with more accurate models of thinking and behavior or providing a transparent explanation of "the prudential, economic, political, or religious reasons for retaining a less accurate and outdated view" (Kang & Banaji 2006, p. 1065). The extent of the pressure to be more behaviorally realistic depends on numerous factors, such as the strength of the scientific consensus regarding the emergent model, the size of the gap between the new model and old assumptions, and the consequences of both action

and omission. This call for behavioral realism is consistent with the increased attention to behavioral economics in legal discourse and calls for a more comprehensive and situation-sensitive psychological portrait of human action, e.g., Hanson's "critical realism" (Hanson & Yosifon 2003).

Within the symposium, Greenwald & Krieger (2006) succinctly summarized the science underlying implicit bias. Krieger & Fiske (2006) discussed how disparate treatment doctrine under Title VII of the 1964 Civil Rights Act might be made more behaviorally realistic, specifically by challenging the "honest belief" rule (that an employer should escape liability if it honestly believed in the nondiscriminatory reasons it provided) and the "same actor" inference (that if the person who fired an employee was also the person who had hired that employee, there would be an inference of no discrimination).

Kang & Banaji (2006) revised various affirmative action arguments into new "fair measures." In particular, they reframed certain affirmative action programs, not as reparations for the past or ways to achieve general future social benefits, but as specific techniques to counter present implicit bias. One such technique would be to deploy "debiasing agents" who are specifically selected for their countertypical attributes that decrease implicit bias instead of standard "role models," which have a dubious constitutional and Title VII status.

Blasi & Jost (2006) provided the first systematic law review introduction to system justification theory, which analyzes the motivation to defend and justify the extant social order, and examined its implications for effective legal and social justice advocacy. Jolls & Sunstein (2006) supplied a theoretical description of ways that the law might respond to implicit bias, for example through alternative methods of debiasing. Finally, Banks et al. (2006) turned to criminal law and examined issues of profiling, sentencing, and shooting in light of implicit bias to expose the lack of consensus on what racial equality actually entails.

The aforementioned works provide a sound introduction to the emergent legal literature on implicit bias and the law. In addition, dozens of articles have referred to the idea of implicit bias generally and the IAT in particular. Most make only passing reference, but some have engaged more substantially with the science and its legal implications. Noteworthy examples include discussions of petit (Page 2005) and grand jury (Teshima 2006) selection, implicit bias of capital defense attorneys (Eisenberg & Johnson 2004), health-care delivery (Shin 2002), and employment discrimination (Bagenstos 2006, Poirier 2003).

The articles mentioned so far largely embrace the implicit bias research as credible, reliable, and illuminating. Others have been more skeptical. For instance, Rachlinski and colleagues (2007) question whether laboratory findings of implicit bias predict behaviors in the real world that surpass some minimum threshold of moral or legal significance. Judges completed an IAT measure of racial attitudes and then sentenced fictional defendants based on a paper profile. In this study, judges' ISCs were uncorrelated with their judgments. Null results such as this one can appear for two reasons—either because no effect of implicit bias predicting behavior actually exists or because such an effect exists but fails to be detected because of an underpowered or otherwise flawed test (the latter is referred to as a Type 2 error in statistics). The relatively small sample of judges ($N = 70$) suggests that their null result may be a Type 2 error, as does the diverse nature of their sample (37 white, 30 black, and 3 Latino), the different distributions of bias among black and white Americans, and subjects' different priming conditions, which could have added systematic variability that masked a bias-behavior relationship.

The Poehlman et al. (2007) meta-analysis provides the best response to this predictive validity concern. Rachlinski et al. (2007) are correct in noting that the Poehlman paper defined behavior expansively as "any measure of

a physical action, judgment, decision or physiological reaction." Awkward body language is hardly in and of itself actionable. That said, such intermediate behaviors should not be dismissed as unimportant, for they can contribute to the final decision-making process and influence marginal cases. Summed over large populations engaged in daily interactions and evaluations, the aggregate impact on individuals and groups may be substantial. Additionally, studies in the meta-analysis specifically measured final or ultimate decisions, including hiring recommendations (Rudman & Glick 2001) and doctors' treatment recommendations (Green et al. 2007). Finally, other studies have provided further support in additional contexts, such as job evaluations (Ziegert & Hanges 2005). Of course, further research will help clarify the behavioral consequences of implicit bias.

In contrast to Rachlinski's reasonable calls for caution, Mitchell & Tetlock (2007) label the large body of work on implicit bias to be pseudoscience created by ideologues in order to warp public policy. In our own work (J. Kang, K.A. Lane & M.R. Banaji, manuscript in preparation), we suggest that this accusation of junk science should be understood as predictable political backlash, regrettably laced with ad hominem and strawperson excess. Indeed, Bagenstos (2007) characterized their appraisal as "based not on any 'scientific' ground, but on normative assumptions . . . rest[ed] on a very narrow view, based on notions of individual fault, that the law should prohibit only discrimination that results from irrational animus."

Judicial Opinions

We have reviewed how the science has moved from psychology journals into law reviews, but what about the next move into the case law? Just as various theories of unconscious bias have long been pondered in law reviews, many judicial opinions have noted the existence and potential legal significance of unconscious forms of bias. Indeed, so many such

opinions make some mention of unconscious or implicit bias that it would be impossible to try to list them all.

In both concurring and dissenting opinions, the Supreme Court has acknowledged the potential for implicit bias to impede justice. For example, in her dissent in *Adarand Constructors, Inc. v. Peña* (1995), Justice Ginsburg noted that "[b]ias both conscious and unconscious, reflecting traditional and unexamined habits of thought, keeps up barriers that must come down if equal opportunity and nondiscrimination are ever genuinely to become this country's law and practice." She echoed these sentiments in the University of Michigan affirmative action cases, quoting these precise words in her dissent in *Gratz v. Bollinger* (2003), and writing in her concurring opinion in *Grutter v. Bollinger* (2003), that "[i]t is well documented that conscious and unconscious race bias, even rank discrimination based on race, remain alive in our land, impeding realization of our highest values and ideals."

More specifically, the Court has speculated that implicit bias may affect the perceptions of participants in the legal system: In *Batson v. Kentucky* (1989), Justice Marshall suggested in his concurring opinion that "[a] prosecutor's own conscious or unconscious racism may lead him easily to the conclusion that a prospective black juror is 'sullen,' or 'distant,' a characterization that would not have come to his mind if a white juror had acted identically." Justice O'Connor voiced a similar concern in her dissent in *Georgia v. McCollum* (1992), noting "[i]t is by now clear that conscious and unconscious racism can affect the way white jurors perceive minority defendants and the facts presented at their trials, perhaps determining the verdict of guilt or innocence." She then offered a possible remedy, suggesting that "[u]sing peremptory challenges to secure minority representation on the jury may help to overcome such racial bias, for there is substantial reason to believe that the distorting influence of race is minimized on a racially mixed jury."

In *Price Waterhouse v. Hopkins* (1989), Justice Brennan, writing for the plurality, suggested that lack of awareness of bias does not excuse the bias or outcomes resulting from it: "unwitting or ingrained bias is no less injurious or worthy of eradication than blatant or calculated discrimination"; "the fact that some or all of the partners at Price Waterhouse may have been unaware of that motivation, even within themselves, neither alters the fact of its existence nor excuses it."

Among appellate court opinions that address implicit or unconscious bias, the most interesting ones consider whether Title VII can account for such cognitive processes. Take, for example, *Thomas v. Eastman Kodak Co.* (1999), which clarified that "[t]he ultimate question is whether the employee has been treated disparately 'because of race.' This is so regardless of whether the employer consciously intended to base the evaluations on race, or simply did so because of unthinking stereotypes or bias."

If we raise the threshold of relevance and discuss only those cases that directly rely on implicit social cognition research for a factual finding or legal conclusion, few cases warrant mention. Two cases discuss implicit bias research specifically, but in dicta and toward opposite conclusions about its significance. In *Chinn v. Runnels* (2004), a habeas corpus petitioner challenged the San Francisco Superior Court's grand jury selection process for never having selected a foreperson of Chinese, Filipino, or Latino descent for 36 years. Given the very deferential standard of review, the federal court accepted the California appellate court's ruling that the government had successfully rebutted the prima facie case of discrimination. However, in the conclusion, the federal court wrote extensively to explain why under a de novo standard of review there could have been a different result. The court specifically cited a "growing body of social science [that] recognizes the pervasiveness of unconscious racial and ethnic stereotyping and group bias." Although many of the citations to the academic literature were to early, important critical race theory literature (e.g., Lawrence 1987), other references were made to the more social cognitive literature (citations made to Bargh 1989, Krieger 1995).

United States v. Taveras (2006), provides dicta that go the other way—minimizing the significance of implicit bias. At issue was the admissibility of potentially prejudicial evidence in the penalty phase of a murder conviction. In dictum, Judge Jack Weinstein opined that

> Judges, who are often exposed to the nastier elements of human behavior on a regular basis, are likely to be better able than laypersons to control the effects of prejudicial information on their decisions. cf. Jeffrey J. Rachlinski, et al. Does Unconscious Bias Affect Trial Judges? (2006) (unpublished manuscript, on file with court) (empirical study of implicit associations suggests that judges, while still susceptible to unconscious biases, are able to set them aside when rendering judgment even when primed with information designed to elicit negative reactions).

The Rachlinski article cited above cautioned about predictive validity. From the lack of correlation between judges' ISCs and sentencing decisions, Judge Weinstein inferred a certain judicial objectivity, driven by prior and regular exposure to the "nastier elements of human behavior." As anybody who works with correlational data knows, it is dangerous to interpret a null result as showing a lack of relationship, especially given the large demand characteristics cueing the judges as to how they should respond on the sentencing exercise and given the low statistical power to detect an effect. This null result should be read in light of the Poehlman et al. (2007) meta-analysis showing a systematic relationship between implicit bias and behavior. Of course, Judge Weinstein did not follow such a reading in *Taveras*. It is noteworthy, and perhaps not entirely surprising, that one of the first references to the literature in a published

judicial opinion proposed that judges are immune to implicit bias.

Finally, in *Farrakhan v. Gregoire* (2006) implicit bias expert testimony apparently influenced an important judicial finding of discrimination—although that finding was nevertheless insufficient for plaintiff's victory. The case addressed a §2 Voting Rights Act (VRA) challenge to Washington state's felon disenfranchisement statute. Surprisingly, the court found "compelling evidence of racial discrimination and bias in Washington's criminal justice system." This finding was not based "solely on statistics" that showed racial disparities in criminal law enforcement. Instead, it was based on expert testimony that included substantial discussion of both structural/institutional factors and, more important to this discussion, the potential impact of implicit bias. One of the two principal expert reports relied upon by the court contained an extensive discussion of implicit bias. In addition, another source described as helping to "bolster the Court's conclusion" included the expert testimony of Anthony Greenwald, inventor of the IAT, who attached a draft of his submission to the Behavioral Realism symposium.

This finding of racial discrimination was, however, insufficient to find a VRA violation under a "totality of circumstances" test. The court heavily emphasized that the state of Washington did not have any history of official discrimination that undermined minority voting rights. Emphasizing this factor seems to miss the point of how implicit bias functions. Nonetheless, *Farrakhan* demonstrates how implicit social cognitive explanations can produce judicial findings of discrimination that probably would not have been found by statistical dispari-

ties alone. As of this writing, the case is on appeal.

CONCLUSION

In a recent commentary, Banaji (2007) wrote that "[t]o be intelligent means many things of course.... [I]ntelligence is knowing how to weigh the evidence that flies in the face of steadfast assumptions. It means to know when causality can be inferred and not, to know when the weight of correlational evidence must be taken seriously, to know that a replication is worth much more than a single demonstration, to know that when new methods divulge strange truths about us and our brethren, it may be the theory that has to go. The moral obligation to be intelligent requires that we keep abreast of discoveries that require old views to be bagged and put out on the curb for recycling—every week."

The eager engagement of legal scholars and lawyers in the discoveries about implicit social cognition comes from the challenge they pose to existing assumptions about human nature that currently reside in the law. It is not for this review to take a position on how such evidence should inform the law. However, this review should reveal the robustness of the evidence that much of human cognition can and does occur without introspective access, that such processes nevertheless influence and guide decision making, that the costs incurred by individuals and social groups are less at the hands of the malign and more likely to come from the unaware and uncontrolled mental acts of ordinary people. The law will need to include these discoveries about how the mind (really) works to be true to Erskine's idea that intelligence about such matters, not just meaning well, is the virtue.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

Adarand Constructors, Inc. v. Peña, 515 U.S. 200 (1995)

Amodio DM, Devine PG. 2006. Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *J. Pers. Soc. Psychol*. 91:652–61

Amodio DM, Harmon-Jones E, Devine PG. 2003. Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *J. Pers. Soc. Psychol*. 84:738–53

Armour J. 1995. Stereotypes and prejudice: helping legal decisionmakers break the prejudice habit. *Calif. Law Rev*. 83:733–72

Ashburn-Nardo L, Knowles ML, Monteith MJ. 2003. Black Americans' implicit racial associations and their implications for intergroup judgment. *Soc. Cogn*. 21:61–87

Ashburn-Nardo L, Voils CI, Monteith MJ. 2001. Implicit associations as the seeds of intergroup bias: How easily do they take root? *J. Pers. Soc. Psychol*. 81:789–99

Bagenstos SR. 2006. The structural turn and the limits of antidiscrimination law. *Calif. Law Rev*. 94:1–47

Bagenstos SR. 2007. Implicit bias, "science," and antidiscrimination law. *Harvard Law Policy Rev.* 1:477–93

Banaji MR. 2001. Implicit attitudes can be measured. In *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, ed. HL Roediger, JS Nairne, I Neath, A Surprenant, pp. 117–50. Washington, DC: Am. Psychol. Assoc.

Banaji MR. 2007. Foreword. In *Beyond Common Sense: Psychological Science in the Courtroom*, ed. E Borgida, ST Fiske. Oxford: Blackwell. In press

Banaji MR, Bazerman MH, Chugh D. 2003. How (un)ethical are you? *Harvard Bus. Rev*. 81:56–64

Banaji MR, Bhaskar R. 2000. Implicit stereotypes and memory: the bounded rationality of social beliefs. In *Memory, Brain, and Belief*, ed. DL Schacter, E Scarry, pp. 139–75. Cambridge, MA: Harvard Univ. Press

Banaji MR, Greenwald AG. 1995. Implicit gender stereotyping in judgments of fame. *J. Pers. Soc. Psychol*. 68:181–98

Banks RR, Eberhardt JL, Ross L. 2006. Discrimination and implicit racial bias in a racially unequal society. *Calif. Law Rev*. 94:1169–90

Barden J, Maddux WW, Petty RE, Brewer MB. 2004. Contextual moderation of racial bias: the impact of social roles on controlled and automatically activated attitudes. *J. Pers. Soc. Psychol*. 87:5–22

Bargh JA. 1989. Conditional automaticity: varieties of automatic influence in social perception and cognition. In *Unintended Thought*, ed. JS Uleman, JA Bargh, pp. 3–51. New York: Guilford

Batson v. Kentucky, 476 U.S. 79 (1989)

Bayern SJ, ed. 2006. Symposium on behavioral realism. *Calif. Law Rev*. Vol. 94, No. 4

Bazerman MH, Chugh D, Banaji MR. 2005. When good people (seem to) negotiate in bad faith. *Negotiation* 8(10):3–5

Bessenoff GR, Sherman JW. 2000. Automatic and controlled components of prejudice toward fat people: evaluation vs stereotype activation. *Soc. Cogn*. 18:329–53

Blair IV. 2002. The malleability of automatic stereotypes and prejudice. *Pers. Soc. Psychol. Rev*. 6:242–61

Blair IV, Ma JE, Lenton AP. 2001. Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *J. Pers. Soc. Psychol*. 81:828–41

Blascovich J, Mendes WB, Hunter SB, Lickel B, Kowai-Bell N. 2001. Perceiver threat in social interactions with stigmatized others. *J. Pers. Soc. Psychol*. 80:253–67

Blasi G. 2002. Advocacy against the stereotype: lessons from cognitive social psychology. *UCLA Law Rev*. 49:1241–81

Blasi G, Jost JT. 2006. System justification theory and research: implications for law, legal advocacy, and social justice. *Calif. Law Rev*. 94:1119–68

Bosson JK, Swann WBJ, Pennebaker JW. 2000. Stalking the perfect measure of implicit self-esteem: the blind men and the elephant revisited? *J. Pers. Soc. Psychol*. 79:631–43

Brendl CM, Markman AB, Messner C. 2005. Indirectly measuring evaluations of several attitude objects in relation to a neutral reference point. *J. Exp. Soc. Psychol*. 41:346–68

Cadinu MR, Rothbart M. 1996. Self-anchoring and differentiation processes in the minimal group setting. *J. Pers. Soc. Psychol*. 70:661–77

Carney DR, Banaji MR. 2007. Social cognition and social neuroscience. In *Cognition*, ed. M Tarr. MA: Sinauer. In press

Chinn v. Runnels, 343 F.Supp.2d 891 (N.D. Ca. 2004)

Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Earlbaum. 2nd ed.

Correll J, Park B, Judd CM, Wittenbrink B. 2002. The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. *J. Pers. Soc. Psychol*. 83:1314–29

Cunningham WA, Johnson MK, Gatenby JC, Gore JC, Banaji MR. 2003. Neural components of social evaluation. *J. Pers. Soc. Psychol*. 85:639–49

Cunningham WA, Johnson MK, Raye CL, Gatenby JC, Gore JC, Banaji MR. 2004a. Separable neural components in the processing of black and white faces. *Psychol. Sci*. 15:806–13

Cunningham WA, Nezlek JB, Banaji MR. 2004b. Implicit and explicit ethnocentrism: revisiting the ideologies of prejudice. *Pers. Soc. Psychol. Bull*. 30:1332–46

Cunningham WA, Preacher KJ, Banaji MR. 2001. Implicit attitude measures: consistency, stability, and convergent validity. *Psychol. Sci*. 121:163–70

Dasgupta N, Asgari S. 2004. Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *J. Exp. Soc. Psychol*. 40:642–58

Dasgupta N, Greenwald AG. 2001. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol*. 81:800–14

Dasgupta N, Rivera LM. 2006. From automatic antigay prejudice to behavior: the moderating role of conscious beliefs about gender and behavioral control. *J. Pers. Soc. Psychol*. 91:268–80

De Houwer J. 2003. The extrinsic affective Simon task. *Exp. Psychol*. 50:77–85

de Jong PJ, van den Hout MA, Rietbroek H, Huijding J. 2003. Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cogn. Emot*. 17:521–45

Devine PG, Plant EA, Amodio DM, Harmon-Jones E, Vance SL. 2002. The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *J. Pers. Soc. Psychol*. 82:835–48

Devos T, Banaji MR. 2005. American = white? *J. Pers. Soc. Psychol*. 88:447–66

Dovidio JF, Kawakami K, Gaertner SL. 2002. Implicit and explicit prejudice and interracial interaction. *J. Pers. Soc. Psychol*. 82:62–68

Egloff B, Schmukle SC. 2002. Predictive validity of an Implicit Association Test for assessing anxiety. *J. Pers. Soc. Psychol*. 83:1441–55

Eisenberg T, Johnson SL. 2004. Implicit racial attitudes of death penalty lawyers. *DePaul Law Rev*. 53:1539–56

Farrakhan v. Gregoire, 2006 U.S. Dist. LEXIS 45987 (E.D. Wash 2006)

Fazio RH, Chen J, McDonel EC, Sherman SJ. 1982. Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *J. Exp. Soc. Psychol*. 18:339–57

Fazio RH, Jackson JR, Dunton BC, Williams CJ. 1995. Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J. Pers. Soc. Psychol*. 69:1013–27

Fazio RH, Olson MA. 2003. Implicit measures in social cognition research: their meaning and use. *Annu. Rev. Psychol*. 54:297–327

Fazio RH, Towles-Schwen T. 1999. The MODE model of attitude-behavior processes. In *Dual-Process Theories in Social Psychology*, ed. S Chaiken, Y Trope, pp. 97–116. New York: Guilford

Gaertner SL, Mann J, Murrell A, Dovidio JF. 1989. Reducing intergroup bias: the benefits of recategorization. *J. Pers. Soc. Psychol*. 57:239–49

Gawronski B, Geschke D, Banse R. 2003. Implicit bias in impression formation: associations influence the construal of individuating information. *Eur. J. Soc. Psychol*. 33:573–89

Georgia v. McCollum, 505 U.S. 42 (1992)

Gratz v. Bollinger, 539 U.S. 244 (2003)

Gray NS, Brown AS, MacCulloch MJ, Smith J, Snowden RJ. 2005. An implicit test of the associations between children and sex in pedophiles. *J. Abnorm. Psychol*. 114:304–8

Gray NS, MacCulloch MJ, Smith J, Morris M, Snowden RJ. 2003. Violence viewed by psychopathic murderers. *Nature* 423:497–98

Green AR, Carney DR, Pallin DJ, Ngo LH, Raymond KL, et al. 2007. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *J. Gen. Intern. Med*. 22:1231–38

Greenwald AG, Banaji MR. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev*. 102:4–27

Greenwald AG, Farnham SD. 2000. Using the Implicit Association Test to measure self-esteem and self-concept. *J. Pers. Soc. Psychol*. 79:1022–38

Greenwald AG, Krieger LH. 2006. Implicit bias: scientific foundations. *Calif. Law Rev*. 94:945–67

Greenwald AG, McGhee DE, Schwartz JLK. 1998. Measuring individual differences in implicit cognition: the Implicit Association Test. *J. Pers. Soc. Psychol*. 74:1464–80

Greenwald AG, Nosek BA, Banaji MR. 2003a. Understanding and using the Implicit Association Test: I: An improved scoring algorithm. *J. Pers. Soc. Psychol*. 85:197–216

Greenwald AG, Oakes MA, Hoffman HG. 2003b. Targets of discrimination: effects of race on responses to weapons holders. *J. Exp. Soc. Psychol*. 39:399–405

Grutter v. Bollinger, 539 U.S. 306 (2003)

Hanson J, Yosifon D. 2003. The situation: an introduction to the situational character, critical realism, power economics, and deep capture. *Univ. Penn. Law Rev*. 152:129–346

Hofmann W, Gawronski B, Gschwendner T, Le H, Schmitt M. 2005. A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Pers. Soc. Psychol. Bull.* 31:1369–85

Hugenberg K, Bodenhausen GV. 2003. Facing prejudice: implicit prejudice and the perception of facial threat. *Psychol. Sci.* 14:640–43

Jolls C, Sunstein C. 2006. The law of implicit bias. *Calif. Law Rev.* 94:969–96

Jost JT, Banaji MR. 1994. The role of stereotyping in system-justification and the production of false consciousness. *Br. J. Soc. Psychol.* 33:1–27

Jost JT, Banaji MR, Nosek BA. 2004. A decade of system justification theory: accumulated evidence of conscious and unconscious bolstering of the status quo. *Polit. Psychol.* 25:881–920

Jost JT, Pelham BW, Carvallo MR. 2002. Non-conscious forms of system justification: implicit and behavioral preferences for higher status groups. *J. Exp. Soc. Psychol.* 38:586–602

Kang J. 2005. Trojan horses of race. *Harvard Law Rev.* 118:1491–593

Kang J, Banaji MR. 2006. Fair measures: a behavioral realist revision of affirmative action. *Calif. Law Rev.* 94:1063–118

Karpinski A, Hilton JL. 2001. Attitudes and the Implicit Association Test. *J. Pers. Soc. Psychol.* 81:774–88

Kawakami K, Dovidio JF, Moll J, Hermsen S, Russin A. 2000. Just say no (to stereotyping): effects of training in the negation of stereotypic associations on stereotype activation. *J. Pers. Soc. Psychol.* 78:871–88

Krieger LH. 1995. The content of our categories: a cognitive bias approach to discrimination and equal employment opportunity. *Stanford Law Rev.* 47:1161–248

Krieger LH. 1998. Civil rights perestroika: intergroup relations after affirmative action. *Calif. Law Rev.* 86:1276–329

Krieger LH, Fiske ST. 2006. Behavioral realism in employment discrimination law: implicit bias and disparate treatment. *Calif. Law Rev.* 94:997–1062

Kühnen U, Schiessl M, Bauer N, Paulig N, Pöhlmann C, Schmidthals K. 2001. How robust is the IAT? Measuring and manipulating implicit attitudes of East- and West-Germans. *Z. Exp. Psychol.* 48:135–44

Lane KA, Banaji MR, Nosek BA, Greenwald AG. 2007. Understanding and using the Implicit Association Test: IV: what we know (so far) about the method. In *Implicit Measures of Attitudes*, ed. B Wittenbrink, N Schwarz, pp. 59–102. New York: Guilford

Lane KA, Mitchell JP, Banaji MR. 2005. Me and my group: cultural status can disrupt cognitive consistency. *Soc. Cogn.* 23:353–86

Lawrence CRI. 1987. The id, the ego, and equal protection: reckoning with unconscious racism. *Stanford Law Rev.* 39:317–88

Lemm KM. 2006. Positive associations among interpersonal contact, motivation, and implicit and explicit attitudes toward gay men. *J. Homosex.* 51:79–99

Livingston RW. 2002. The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *J. Exp. Soc. Psychol.* 38:405–13

McConnell AR, Leibold JM. 2001. Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *J. Exp. Soc. Psychol.* 37:435–42

Mitchell G, Tetlock PE. 2007. Antidiscrimination law and the perils of mindreading. *Ohio State Law J.* 67:1023

Mitchell JP, Macrae CN, Banaji MR. 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50:655–63

Mitchell JP, Nosek BA, Banaji MR. 2003. Contextual variations in implicit evaluation. *J. Exp. Psychol. Gen.* 132:455–69

Nock MK, Banaji MR. 2007a. Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *J. Consult. Clin. Psychol.* In press

Nock MK, Banaji MR. 2007b. Assessment of self-injurious thoughts using a behavioral test. *Am. J. Psychiatry* 164:820–32

Nosek BA. 2005. Moderators of the relationship between implicit and explicit evaluation. *J. Exp. Psychol. Gen.* 134:565–84

Nosek BA, Banaji MR. 2001. The go/no-go association task. *Soc. Cogn.* 19:625–66

Nosek BA, Banaji M, Greenwald AG. 2002. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dyn.* 6:101–15

Nosek BA, Greenwald AG, Banaji MR. 2006. The Implicit Association Test at age 7: a methodological and conceptual review. In *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*, ed. JA Bargh, pp. 265–92. New York: Psychology Press

Nosek BA, Hansen JJ. 2007. The associations in our heads belong to us: searching for attitudes and knowledge in implicit evaluation. *Cogn. Emot.* In press

Nosek BA, Smyth FL, Hansen JJ, Devos T, Linder NM, et al. 2007. Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* In press

Olson MA, Fazio RH. 2004. Trait inferences as a function of automatically activated racial attitudes and motivation to control prejudiced reactions. *Basic Appl. Soc. Psychol.* 26:1–11

Olson MA, Fazio RH. 2006. Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Pers. Soc. Psychol. Bull.* 32:421–33

Ottaway SA, Hayden DC, Oakes MA. 2001. Implicit attitudes and racism: effects of word familiarity and frequency on the Implicit Association Test. *Soc. Cogn.* 19:97–144

Page A. 2005. *Batson*'s blind-spot: unconscious stereotyping and the peremptory challenge. *Boston Univ. Law Rev.* 85:155–262

Payne BK. 2001. Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *J. Pers. Soc. Psychol.* 81:181–92

Payne BK, Cheng CM, Govorun O, Stewart BD. 2005. An inkblot for attitudes: affect misattribution as implicit measurement. *J. Pers. Soc. Psychol.* 89:277–93

Perdue CW, Dovidio JF, Gurtman MB, Tyler RB. 1990. Us and them: social categorization and the process of intergroup bias. *J. Pers. Soc. Psychol.* 59:475–86

Peruche BM, Plant EA. 2006. Racial bias in perceptions of athleticism: the role of motivation in the elimination of bias. *Soc. Cogn.* 24:438–52

Phelps EA. 2006. Emotion and cognition: insights from studies of the human amygdala. *Annu. Rev. Psychol.* 57:27–53

Phelps EA, O'Connor KJ, Cunningham WA, Funayama ES, Gatenby JC, et al. 2000. Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* 12:729–38

Plant EA, Peruche BM. 2005. The consequences of race for police officers' responses to criminal suspects. *Psychol. Sci.* 16:180–83

Plant EA, Peruche BM, Butz DA. 2005. Eliminating automatic racial bias: making race non-diagnostic for responses to criminal suspects. *J. Exp. Soc. Psychol.* 41:141–56

Poehlman TA, Uhlmann EL, Greenwald AG, Banaji MR. 2007. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Manuscript under review

Poirier MR. 2003. Is cognitive bias at work a dangerous condition on land? *Empl. Rights Employ. Policy J.* 7:459–94

Pollard DA. 1999. Unconscious bias and self-critical analysis: the case for a qualified evidentiary equal employment opportunity privilege. *Wash. Law Rev.* 74:913–1031

Pratto F, Shih M. 2000. Social dominance orientation and group context in implicit group prejudice. *Psychol. Sci*. 11:515–18

Price Waterhouse v. Hopkins, 490 U.S. 228 (1989)

Rachlinski JJ, Johnson S, Wistrich AJ, Guthrie C. 2007. Does unconscious bias affect trial judges? Work. Pap., Cornell Univ., Ithaca, NY

Richeson JA, Ambady N. 2001. Who's in charge? Effects of situational roles on automatic gender bias. *Sex Roles* 44:493–512

Richeson JA, Shelton JN. 2003. When prejudice does not pay: effects of interracial contact on executive function. *Psychol. Sci*. 14:287–90

Rudman LA, Feinberg J, Fairchild K. 2002. Minority members' implicit attitudes: automatic ingroup bias as a function of group status. *Soc. Cogn*. 20:294–320

Rudman LA, Glick P. 2001. Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* 57:743–62

Rudman LA, Kilianski SE. 2000. Implicit and explicit attitudes toward female authority. *Pers. Soc. Psychol. Bull*. 26:1315–28

Rudman LA, Lee MR. 2002. Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Process. Intergroup Relat*. 5:133–50

Saujani RM. 2003. The Implicit Association Test: a measure of unconscious racism in legislative decision-making. *Mich. J. Race Law* 8:395–424

Sekaquaptewa D, Espinoza P, Thompson M, Vargas P, von Hippel W. 2003. Stereotypic explanatory bias: implicit stereotyping as a predictor of discrimination. *J. Exp. Soc. Psychol*. 39:75–82

Shih M, Pittinsky TL, Ambady N. 1999. Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychol. Sci*. 10:80–83

Shin MS. 2002. Redressing wounds: finding a legal framework to remedy racial disparities in medical care. *Calif. Law Rev*. 90:2047–100

Shutts K, Olson KA, Spelke ES, Banaji MR. 2007. Implicit attitudes in postapartheid South Africa. Work. Pap., Harvard Univ., Cambridge, MA

Sinclair L, Kunda Z. 1999. Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes. *J. Pers. Soc. Psychol*. 77:885–904

Sinclair S, Lowery BS, Hardin CD, Colangelo A. 2005. Social tuning of automatic racial attitudes: the role of affiliative motivation. *J. Pers. Soc. Psychol*. 89:583–92

Snowden RJ, Gray NS, Smith J, Morris M, MacCulloch MJ. 2004. Implicit affective associations to violence in psychopathic murderers. *J. Forensic Psychiatry Psychol*. 15:620–41

Swanson JE, Rudman LA, Greenwald AG. 2001. Using the Implicit Association Test to investigate attitude-behaviour consistency for stigmatised behaviour. *Cogn. Emot*. 15:207–30

Tajfel H, Billig MG, Bundy RP, Flament C. 1971. Social categorization and intergroup behaviour. *Eur. J. Soc. Psychol*. 1:149–78

Tajfel H, Turner JC. 1986. The social identity theory of intergroup behavior. In *Psychology of Intergroup Relations*, ed. S Worchel, LW Austin, pp. 7–24. Chicago: Nelson-Hall

Teachman BA, Woody SR. 2003. Automatic processing in spider phobia: implicit fear associations over the course of treatment. *J. Abnorm. Psychol*. 112:100–9

Teshima DS. 2006. A "hardy handshake sort of guy": the model minority and implicit bias about Asian Americans in *Chin v. Runnels*. *Asian Pac. Am. Law J*. 11:122–41

Thomas v. Eastman Kodak Co., 183 F.3d 38 (1st Cir. 1999)

Towles-Schwen T, Fazio RH. 2003. Choosing social situations: the relation between automatically activated racial attitudes and anticipated comfort interacting with African Americans. *Pers. Soc. Psychol. Bull*. 29:170–82

United States v. Taveras, 424 F.Supp.2d 446 (E.D. N.Y. 2006)

Vanman EJ, Saltz JL, Nathan LR, Warren JA. 2004. Racial discrimination by low-prejudiced whites. Facial movements as implicit measures of attitudes related to behavior. *Psychol. Sci.* 15:711–14

Village of Arlington Heights v. Metropolitan Housing Development Authority, 429 U.S. 252 (1977)

von Hippel W, Sekaquaptewa D, Vargas P. 1997. The linguistic intergroup bias as an implicit indicator of prejudice. *J. Exp. Soc. Psychol.* 33:490–509

Washington v. Davis, 426 U.S. 229 (1976)

Wilson TD, Lindsey S, Schooler TY. 2000. A model of dual attitudes. *Psychol. Rev.* 107:101–26

Wittenbrink B. 2007. Measuring attitudes through priming. In *Implicit Measures of Attitudes*, ed. B Wittenbrink, N Schwarz, pp. 17–58. New York: Guilford

Zajonc R. 1980. Feeling and thinking: preferences need no inferences. *Am. Psychol.* 35:151–75

Ziegert JC, Hanges PJ. 2005. Employment discrimination: the role of implicit attitudes, motivation, and a climate for racial bias. *J. Appl. Psychol.* 90:553–62

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Law and Social Science* articles may
be found at http://lawsocsci.annualreviews.org