Helping Courts Address Implicit Bias

Frequently Asked Questions*

NCSC National Center for State Courts www.ncsc.org

Race & Ethnic Fairness in the Courts

1) What is Implicit Bias?

Unlike *explicit bias* (which reflects the attitudes or beliefs that one endorses at a conscious level), *implicit bias* is the bias in judgment and/or behavior that results from subtle cognitive processes (e.g., implicit attitudes and implicit stereotypes) that often operate at a level below conscious awareness and without intentional control. The underlying implicit attitudes and stereotypes responsible for implicit bias are those beliefs or simple associations that a person makes between an object and its evaluation that "...are automatically activated by the mere presence (actual or symbolic) of the attitude object" (**Dovidio, Gaertner, Kawakami, & Hudson, 2002**, p. 94; also **Banaji & Heiphetz, 2010**). Although automatic, implicit biases are not completely inflexible: They are malleable to some degree and manifest in ways that are responsive to the perceiver's motives and environment (**Blair, 2002**).

Implicit bias research developed from the study of attitudes. Scientists realized long ago that simply asking people to report their attitudes was a flawed approach; people may not wish or may not be able to accurately do so. This is because people are often unwilling to provide responses perceived as socially undesirable and therefore tend to report what they think their attitudes *should* be rather than what they know them to be. More complicated still, people may not even be consciously aware that they hold biased attitudes. Over the past few decades, scientists have developed new measures to identify these unconscious biases (see FAQ #3: *How is implicit bias measured?*).

*Preparation of this project brief was funded by the Open Society Institute, the State Justice Institute, and the National Center for State Courts. The views expressed are those of the authors and do not necessarily reflect the views of the funding organizations. The document summarizes the National Center for State Courts' project on implicit bias and judicial education. See Casey, Warren, Cheesman, and Elek (2012), available at www.ncsc.org/ibreport for the full report of the project.

2) What do researchers think are the sources of implicit bias?

Although scientists are still working to understand implicit bias, current theory and evidence indicate that it may arise from several possible sources (as listed by **Rudman, 2004**). These interrelated sources include:

Developmental History

Implicit bias can develop over time with the accumulation of personal experience. Personal experiences include not only traditional learning experiences between the self and the target (i.e., classical conditioning; **Olson & Fazio, 2001**), but also social learning experiences (i.e., via observing parents, friends, or influential others; **Greenwald & Banaji, 1995**). For example, implicit biases in children are positively correlated with the implicit biases of their parents; however, consistent with social learning theory (**Bandura, 1997)**, this congruence occurs only between children who identify with their parents and not for children who do not have a positive attachment relationship with their parents (**Sinclair, Dunn, & Lowery, 2005**). Implicit biases can develop relatively quickly through such experiences: Implicit racial bias has been found in children as young as 6 years old, and discrepancies between implicit and explicit attitudes emerge by the age of 10 (**Baron & Banaji, 2006**).

Affective Experience

Implicit bias may develop from a history of personal experiences that connect certain racial groups with fear or other negative affect. Recent developments in the field of cognitive neuroscience demonstrate a link between implicit (but not explicit) racial bias and neural activity in the amygdala, a region in the brain that scientists have associated with emotional learning and fear conditioning. Specifically, White individuals who score highly on measures of implicit racial bias

also react to images of unfamiliar Black faces with stronger amygdala activation (**Phelps, O'Connor, Cunningham, Funayama, Gatenby, Gore, & Banaji, 2000; see also Stanley, Phelps, & Banaji, 2008**). Other researchers have demonstrated a causal relationship between the experience of certain types of emotions and the emergence of implicit bias, showing that inducing people to experience anger or disgust can create implicit bias against newly encountered outgroups (**Dasgupta, DeSteno, Williams, & Hunsinger, 2009**). Another study found that increased exposure to a socially valued Black instructor in the context of a diversity education course decreased participants' implicit bias against Blacks, and that a reduced fear of Blacks – in addition to other affective factors – predicted this attitudinal change (**Rudman, Ashmore, & Gary, 2001**).

Culture

People share a common social understanding of the stereotypes that are pervasive in our culture, and this knowledge can foster implicit bias even if a person does not necessarily endorse the cultural stereotype (**Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995**). One explanation is that people implicitly make associations and evaluations based on cultural knowledge in a way that "may not be available to introspection and may not be wanted or endorsed but is still *attitudinal* because of its potential to influence individual perception, judgment, or action" (**Nosek, 2007**, p. 68 [emphasis added]). Another explanation offered by Nosek (2007) is that responses on implicit measures are easily influenced by cultural knowledge, but that this cultural knowledge does not reflect the respondent's actual attitude (e.g., **Karpinski & Hilton, 2001**).

The Self

People tend to possess consistent and strongly positive attitudes toward themselves, and this positive attitude about the self can transfer very easily to other things, people, and groups that share attributes with the self (for a review, see **Banaji & Heiphetz, 2010**). This transference can occur without conscious awareness; hence, such effects are termed "implicit egotism." For example, people demonstrate a biased preference for new products that resemble their own names (**Brendl, Chattopadhyay, Pelham, & Carvallo, 2005**). They appear to be disproportionately likely to live in locations that reflect their birth date (e.g., people born on February 2nd and residing in the town of Two Rivers, Wisconsin) and to choose careers or marry others with names that resemble their own (e.g., people named Dennis or Denise in dentistry, a marriage between two unrelated Smiths). They are also more attracted than usual to others who have been assigned an allegedly random experimental code number that matches their birth dates and whose alleged surnames share letters with their own surnames (**Pelham, Mirenberg, & Jones, 2002; Jones, Pelham, Carvallo, & Mirenberg, 2004**). Provocative and strange, this research illustrates the impressive automaticity of the human mind and the influence of implicit processes in our daily lives. Fundamental attitudes toward the self may underlie implicit racial bias by facilitating a general tendency to prefer one's ingroup (a group with which one identifies in some way) over outgroups (any group with which one does not affiliate; see **Greenwald, Banaji, Rudman, Farnham, Nosek, & Mellott, 2002**). As **Rudman (2004)** explains, people tend to believe that "If I am good and I am X [*X being any social group with which one identifies*], then X is also good" (p. 137; italicized text added).

3) How is Implicit Bias measured?

Researchers use a number of scientific methods in the measurement of implicit bias (for reviews, see **Fazio & Olson, 2003; Gawronski, 2009; Wittenbrink & Schwarz, 2007**). Although the specific procedures involved in the individual approaches differ widely, implicit measures take on one of the following three general forms:

Computerized Measures

Computerized implicit measures typically gauge the direction and strength of a person's implicit attitudes by assessing their reaction times (i.e., response latencies) when completing a specific computerized task. The exact nature of each task varies, but usually falls into one of two classes of procedures (see **Wittenbrink & Schwarz, 2007**): *sequential priming or response competition*.

Sequential priming procedures. Sequential priming procedures are based on a long history of evidence in the field of cognitive psychology demonstrating that when two concepts are related in memory, the presentation of one of those concepts facilitates the recall or recognition of the other (see **Neely, 1991**). In the context of racial bias, people with a negative implicit racial bias toward Blacks will more quickly and easily respond to concepts associated with the negative stereotype of Blacks than concepts that are not associated with that stereotype. One popular procedure for measuring this phenomenon is the evaluative priming task or "bona-fide pipeline" (**Fazio, Sanbonmatsu, Powell, & Kardes, 1986**). In this task, respondents are briefly presented with a Black or White face immediately before a positive or negative target word appears on the screen. They must then identify, as quickly as possible, the meaning of the presented word as "good" or "bad." In the standard paradigm, respondents with racial bias more quickly identify negative words as "bad" and more slowly identify positive

words as "good" when that word appears immediately after the presentation of a Black face (**Fazio, et al., 1995**). A similar priming procedure, called the Affect Misattribution Procedure (AMP; **Payne, Cheng, Govorun, & Stewart, 2005**), briefly presents respondents with a prime of a Black or White face before viewing a neutral Chinese character they know they must evaluate as more or less visually pleasant than the average Chinese character. These researchers found that individuals' racial attitudes colored their evaluations of the characters, with White respondents reporting more favorable ratings for characters that appeared after White primes than Black primes. This effect emerged even when respondents received a forewarning about the influence of the racial primes on subsequent evaluations.

Response competition procedures. Another approach to implicit attitude measurement emerged from research on interference effects. Specifically, when a target has multiple different meanings (e.g., the word "red" written in blue font), these different meanings can imply competing responses (e.g., color identification as red or blue) in a given task that can slow down the overall performance of the respondent (note that the well-known Stroop effect is one example of interference effects at work; see **Stroop, 1935; MacLeod, 1991**). These implicit measures, called response competition procedures (**Wittenbrink & Schwarz, 2007**), takes advantage of the informational value of interference effects by presenting two competing categorization tasks in a single procedure and measuring response latencies. Thus, unlike the sequential priming procedures discussed above in which shorter response times indicate bias, *longer* response times denote implicit bias when response competition procedures are used. One of the most popular of these types of measures is the Implicit Association Test (IAT; **Greenwald, McGhee, & Schwartz, 1998**). In the IAT, respondents are asked to categorize a sequence of images (as a Black or White face) and words (as either good or bad) by pressing one of two pre-labeled buttons. For example, the respondent may be instructed to press the left

button whenever they see a Black face or whenever a negative word appears, and to press the right button whenever they see a White face or a positive word. Alternatively, they may be informed to press one button when they see a Black face or positive word, and the other button for a White face or negative word. Because of interference effects, individuals who associate "Black" with "bad," for example, will respond much more slowly when "Black" and "good" share the same response button. Related measures include the Go/No-Go Association Task (GNAT; see **Nosek & Banaji, 2001**) and the Extrinsic Affective Simon Task (EAST; see **De Houwer, 2003**).

Paper & Pencil Measures

Several paper & pencil measures of implicit attitudes exist (see **Vargas, Sekaquaptewa, & von Hippel, 2007** for a review). Some of these measures are simply adaptations of existing computerized assessments. Although researchers have primarily focused on developing manual adaptations of the IAT (e.g., **Kitayama & Uchida, 2003; Lemm, Sattler, Khan, Mitchell, & Dahl, 2002**), Vargas and colleagues (2007) suggest that the AMP (see description under "Computerized Measures," above) may be more easily adapted to a paper & pencil format because the procedure does not involve measurement of response time.

Other paper & pencil implicit measures assess memory accessibility. One example is the Word Fragment Completion (WFC) task, in which people are presented with fragments of words (e.g., POLI_E) and are asked to fill in the missing letters. These word fragments, however, can be completed in stereotypic or non-stereotypic ways (e.g., POLITE, POLICE; **Gilbert & Hixon, 1991**). The number of stereotypic word completions in the WFC task has been used as an implicit measure of racial prejudice (e.g., **Son Hing, Li, & Zanna, 2002**).

Finally, two other implicit bias measurement approaches assess attributional processing styles. One such example is the Stereotypic Explanatory Bias (SEB;

Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003), which is the tendency to ascribe the stereotype-consistent behavior of minorities to factors intrinsic to the individual (e.g., trait or dispositional attributions like hard work or talent), but stereotype-inconsistent behavior to extrinsic, situational factors (e.g., the weather, luck). Similarly, the Linguistic Intergroup Bias (LIB; **Maass, Salvi, Arcuri, & Semin, 1989**) is the tendency to describe stereotypic behavior using abstract language (e.g., by ascribing the behavior to a global trait) but non-stereotypic behavior using concrete language (e.g., by describing the behavior as a specific event). By carefully examining the respondent's choice of language or agreement with particular summaries of a behavioral event, researchers have used these tendencies as indicators of implicit prejudice (see **von Hippel, Sekaquaptewa, & Vargas, 1997** and **Sekaquaptewa, et al. 2003**).

Physiological Measures

Psychologists have long expressed interest in determining the physiological correlates of psychological phenomena. Those interested in the study of intergroup attitudes have examined autonomic nervous system responses such as the amount of sweat produced (e.g., **Rankin & Campbell, 1955**), heart rate (e.g., **Shields & Harriman, 1984**), and even small facial muscle movements that are nearly imperceptible to the untrained human eye (e.g., **Vanman, Saltz, Nathan, & Warren, 2004; Mahaffey, Bryan, & Hutchison, 2005**). More recently, neuroscientists have attempted to understand the neural underpinnings of implicit bias (e.g., **Stanley, Phelps, & Banaji, 2008; Cunningham, Johnson, Gatenby, Gore, & Banaji, 2003**). With further technological advances in physiological measurement, researchers will gain greater insight into the connection between psychological and physiological phenomena that could make some physiological techniques invaluable in the measurement and study of implicit bias. Given the current state of the science, however, the following common techniques are appropriate for advancing scientific understanding of implicit bias, but not for the detection of implicit bias (i.e., "diagnosing" implicit bias in an individual).

Common physiological measures used in the study of attitudes (as described more thoroughly in reviews by **Banaji & Heiphetz, 2010; Blascovich & Mendes; 2010**, and **Ito & Cacioppo, 2007**) include:

EDA. The measurement of sweat production is interchangeably referred to as skin conductance response (SCR), galvanic skin response (GSR), and electrodermal activity (EDA). When an individual experiences greater arousal in response to a stimulus, the eccrine glands in the skin (particularly in the hands and feet) excrete more sweat (**Banaji & Heiphetz, 2010**, p. 363). However, sweat production as a response and, therefore, EDA as a measurement tool do not discriminate between positive and negative responses to a stimulus. That is, by itself, EDA provides no information about the valence of the individual's response, but simply detects arousal. For example, as Banaji & Heiphetz (2010) explain, greater EDA in the presence of Black individuals but not White individuals (**Rankin & Campbell, 1955**) indicates only that the respondent reacts more strongly to the Black individual, and not that the reaction is necessarily a negative one.

Cardiovascular responses. Although a number of techniques have been used to measure cardiac and vasomotor responses, the most common measurement is that of heart rate. Like EDA, heart rate is a valence-insensitive measure of autonomic nervous system arousal and therefore cannot be used to distinguish between positive and negative reactions to a stimulus.

EMG. Facial electromyography (EMG) is the measurement of electrical activity associated with facial muscle contractions. With this technique, researchers can detect the presence of muscle movements and measure the amplitude of the response. Unlike some of the earlier measures discussed, however, the facial EMG can be used to assess response valence because different facial muscles are associated with positive and negative reactions. One study found that greater cheek EMG activity towards Whites than Blacks predicted racial

bias in participant selection decisions when evaluating candidates for a teaching fellowship (**Vanman, Saltz, Nathan, & Warren, 2004**). Unlike the IAT, the facial EMG remained unaffected by participants' motivation to control for prejudiced responses, indicating its potential value as a measure of implicit attitudes.

Another physiological measure, the *startle eyeblink response*, relies on similar response mechanisms; however, only highly arousing stimuli evoke a startle response, limiting the utility of this measurement approach.

fMRI. Functional magnetic resonance imaging (fMRI) is a relatively new technique that measures blood flow in the brain. Because increased blood flow in any specific region of the brain signals increased activity in that region, blood flow can be used as a proxy measure for neural activity. In a groundbreaking study, **Phelps, O'Connor, Cunningham, Funayama, Gatenby, Gore, and Banaji (2000)** demonstrated a correlation between the degree of activation in the amygdala region of the brain, as measured by fMRI, and scores on the IAT; moreover, people exhibit greater amygdala activation when processing negative, rather than positive, stimuli (**Cunningham, Johnson, Gatenby, Gore, & Banaji, 2003**). Although other brain areas are involved in social cognitive processes like implicit bias, the amygdala has been extensively studied because it is so important to evaluation and preference development (**Banaji & Heiphetz, 2010**).

ERP. Event-related brain potentials (ERPs) are measurable electrical signals emitted by brain activity (i.e., neural firing) and provide information on the strength and valence of a person's response to a stimulus. Because this technique measures real-time changes (within milliseconds) in neural activity, researchers can correlate individual ERP data with specific temporal events (e.g., changes in brain activity from a baseline measurement after exposure to a photo of a Black man). Several specific components of ERPs (e.g., larger late-positive potentials or LPPs; **Ito, Thompson, & Cacioppo, 2004**) provide information about an individual's responses to others that are related to implicit bias (for more information, see **Ito & Cacioppo, 2007**, pp. 134-138).

4) Does Implicit Bias matter much in the real world?

A recent meta-analysis of 122 research reports found that one implicit measure (the IAT) effectively predicted bias in a range of relevant social behaviors, social judgments, and even physiological responses (r = .274; **Greenwald, Poehlman, Uhlmann, & Banaji, 2009**). Implicit bias can influence a number of professional judgments and actions in the "real world" (see **Jost, Rudman, Blair, Carney, Dasgupta, Glaser & Hardin, 2009**) that may have legal ramifications.

Some particularly relevant examples are:

Police Officers: The Decision to Shoot

Police officers face high-pressure, high-risk decisions in the line of fire. One seminal research report reveals that these rapid decisions are not immune to the effects of implicit biases. Specifically, college participants in this study played a computer game in which they needed to shoot dangerous armed characters as quickly as possible (by pressing a "shoot" button), but decide not to shoot unarmed characters (by pressing a "don't shoot" button). Some of the characters held a gun, like a revolver or pistol, and some of the characters held innocuous objects, like a wallet or cell phone. In addition, half of the characters were White, and half were Black. Study participants more quickly chose to shoot armed Black characters than armed White characters and more quickly chose not to shoot unarmed White characters than unarmed Black characters. They also committed more "false alarm" errors, electing to shoot unarmed Black characters more than unarmed White characters and electing not to shoot armed White characters more than armed Black characters (**Correll, Park, Judd, & Wittenbrink, 2002**). This research was inspired by the 1999 New York City shooting of Guinean immigrant Amadou Diallo: Police officers fired 41 rounds and killed Diallo as he

pulled out a wallet. Other studies produced similar results with police officers and community members, and also showed that training and practice can help to reduce this bias (e.g., **Correll, Park, Judd, Wittenbrink, Sadler, & Keesee, 2007; Plant & Peruche, 2005; Plant, Peruche, & Butz, 2005**).

Physicians: Treatment Decisions

Physicians routinely make crucial decisions about medical care for patients whose lives hang in the balance. In the face of such high stakes, it may be surprising to think that automatic associations can unknowingly bias professional decision-making. One study showed that the implicit racial biases of ER physicians predicted fewer thrombolysis treatment recommendations when the patient was described as Black as opposed to White (**Green, Carney, Pallin, Ngo, Raymond, Iezzoni, & Banaji, 2007**). The implicit racial biases of White physicians also seem to play a role in predicting how positively or negatively Black patients respond to the medical interaction (**Penner, Dovidio, West, Gaertner, Albrecht, Daily, & Markova, 2010**), which might lead to a greater incidence of malpractice lawsuits (cf. **Stelfox, Gandhi, Orav, & Gustafson, 2005**).

Managers: Hiring Decisions

When screening a pool of job candidates, hiring managers must review hundreds if not thousands of resumes of qualified applicants. Studies show that interview and selection decisions reflect bias against minorities (e.g., **Dovidio & Gaertner, 2000; Bertrand & Mullainathan, 2004; Ziegert & Hanges, 2005**). In one such study, hiring managers were three times less likely to call highly qualified Arab job candidates in for an interview compared to equally qualified candidates of the racial majority. Interestingly, the implicit racial bias scores of hiring managers predicted their likelihood of offering callbacks to the Arab job applicants (**Rooth, 2010**).

Helping Courts Address Implicit Bias

Judges and Jurors: Capital Punishment and Sentencing

If implicit biases can affect both the intuitive, split-second decisions of police officers and sway the more deliberate decisions of physicians and hiring managers, it stands to reason that judges and jurors may exhibit similar tendencies. Indeed, one archival study of 600 death-eligible cases in Philadelphia appears to support this possibility. Researchers identified all cases (n=44) in which a Black male defendant was convicted of murdering a White victim and presented a photograph of each defendant to participants, who in turn rated each defendant on how "stereotypically Black" he appeared to be. Stereotypicality of appearance predicted death penalty sentencing outcomes: 57.5% of those judged as more stereotypically Black were sentenced to death, compared to 24.4% of those who were perceived as less stereotypically Black (**Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006**). Eberhardt and colleagues explain this effect in the context of other empirical research (**Eberhardt, Goff, Purdie, & Davies, 2004**) that demonstrates a tendency to implicitly associate Black Americans with crime. Other studies further illustrate racial biases in the context of detain-release decisions, verdicts, and sentencing (e.g., **Gazal-Ayal & Sulitzeanu-Kenan, 2010; Sommers & Ellsworth, 2001**).

Voters and Other Decision-Makers

Other research also shows that implicit racial biases can predict voting intentions and behavior. In one study of 1,057 registered voters, pro-White implicit bias scores predicted reported intent to vote for McCain over Obama a week before the 2008 U.S. Presidential election (**Greenwald, Smith, Sriram, Bar-Anan, & Nosek, 2009**). Another study found that, after controlling for explicit prejudice, voters who were more implicitly prejudiced against Blacks were less likely to vote for Obama and more likely to abstain from the vote or vote for third party candidates (**Payne, Krosnick, Pasek, Lelkes, Akhtar, & Tompson, 2010**). Implicit biases may, in particular, help "tip the scales" for undecided decision-makers (e.g., **Galdi, Arcuri, & Gawronski, 2008**).

5) What are the key criticisms of Implicit Bias research?

The mounting research evidence on the phenomenon of implicit bias may lead to two disconcerting conclusions: (1) People know less about their own mental processes than common sense would suggest, and (2) overt racism may be diminishing, but subtler forms of racism persist. As is often the case with provocative science, this program of research has its proponents and its skeptics. Scholarly debate revolves primarily around the definition and appropriate measurement of implicit bias, and some have questioned the existence of implicit bias as an attitudinal phenomenon.

Some individuals stridently resist the idea of implicit racial prejudice and are vocal about their opposition (e.g., **Mitchell & Tetlock, 2006; Wax & Tetlock, 2005**). These individuals argue that they are "under no obligation to agree when a segment of the psychological research community labels the vast majority of the American population unconsciously prejudiced on the basis of millisecond reaction-time differentials on computerized tests. It is our view that the legal community should require evidence that scores on these tests of unconscious prejudice map in replicable functional forms onto tendencies to discriminate in realistic settings…" and that, because of this and because the IAT is informed by a variety of factors that "cannot plausibly be labeled precursors to discrimination," the IAT does not tap into "100% pure prejudice" (**Mitchell & Tetlock, 2009**).

In response to these criticisms, the proponents of implicit bias argue that the large body of research over several decades and hundreds of neuroscientific, cognitive, and social psychological studies has produced sufficient if not overwhelming evidence to support the existence of the kinds of automatic negative associations referred to as "implicit bias" (for a review and one of many direct responses to the opposing allegations of Tetlock and colleagues,

see **Jost et al., 2009**). An exponentially increasing number of empirical studies demonstrate a relationship between measures of implicit bias and real-world discriminatory behavior (see FAQ #4: *Does Implicit Bias Matter Much in the Real World?*, above). Moreover, attitudes are flexible constructs—not rigid ones—and one's expressed attitude at any given moment is responsive to a variety of relevant and seemingly irrelevant factors. For example, one now-classic study showed that people's judgments of even their own life satisfaction could be influenced by incidental factors such as the weather (i.e., sunny or cloudy) on the day they were surveyed (**Schwarz & Clore, 1983**). Similarly, the expression of implicit bias is sensitive to a range of sometimes subtle moderating factors (e.g., see **Blair, 2002**).

A key component of the implicit bias controversy is the concern that the IAT, specifically, is problematic. Some believe that proponents of the IAT overstate the consequentiality of their research findings (e.g., **Blanton & Jaccard, 2008; Blanton & Jaccard, 2006**), and others argue that although evaluative priming measures may be construed as "automatic evaluations," what exactly the IAT technique measures is debatable (**Fazio & Olson, 2003**). Indeed, the IAT and a popular evaluative priming implicit measure, the bona-fide pipeline, fail to show correspondence with one another even though both are supported by empirical evidence demonstrating correspondence with actual behavior (**Olson & Fazio, 2003**). These researchers and others (e.g., **Karpinski & Hilton, 2001**) argue that the IAT measures not attitudes but extrapersonal associations acquired through the environment, whether those associations are personally endorsed at an attitudinal level or not. In response to this assertion, **Nosek (2007)** argues that regardless of whether these implicit processes are labeled as attitudes or as associations, the effect is still the same: These automatic processes are capable of guiding our thoughts and actions in predictable—and biased—ways.

Helping Courts Address Implicit Bias

Opponents of the IAT have gone on to propose a number of alternative explanations to discount the IAT as a measure of implicit bias, although variation in the interpretation of how the phenomenon is defined may be partly responsible for this scholarly discord. Proponents of the IAT have thus far presented evidence discrediting several, but not all, of these alternative explanations (e.g., **Dasgupta, Greenwald, & Banaji, 2003; Nosek, Greenwald, & Banaji, 2005; Greenwald, Nosek, & Banaji, 2003**; see *Dr. Anthony Greenwald's IAT Page* for a complete listing of relevant research). These disparate views will likely be resolved as the science advances and new methods for the measurement of implicit bias are developed.

6) What can people do to mitigate the effects of Implicit Bias on judgement and behavior?

Once people are made aware of their own implicit biases, they can begin to consider ways in which to address them. Scientists have uncovered several promising implicit bias intervention strategies that may help individuals who strive to be egalitarian:

- Consciously acknowledge group and individual differences (i.e., adopt a multiculturalism approach to egalitarianism rather than a color-blindness strategy in which one tries to ignore these differences)
- Routinely check thought processes and decisions for possible bias (i.e., adopt a thoughtful, deliberative, and self-aware process for inspecting how one's decisions were made)
- Identify sources of stress and reduce them in the decision-making environment
- Identify sources of ambiguity and impose greater structure in the decision-making context
- Institute feedback mechanisms
- Increase exposure to stereotyped group members (e.g., seek out greater contact with the stigmatized group in a positive context)

For more detailed information on promising debiasing strategies, see *Appendix G in Casey, et al. (2012).*

7) Can people eliminate or change an Implicit Bias?

There is a difference between reducing the influence of implicit bias on decisions (see FAQ #6: *What can people do to mitigate the effects of implicit bias on judgment and behavior?*) and reducing implicit bias itself. Although implicit bias is malleable, many "debiasing" strategies seem to only temporarily reduce or shift it. Longer-term change might be possible only through substantial and persistent effort (for a discussion about the conditional limitations of some existing strategies for reducing implicit bias, see **Joy-Gaba & Nosek, 2010**).

If applied long-term, people may be able to reduce or eliminate implicit bias by modifying their underlying implicit attitudes. Generally, increased contact with or exposure to a stigmatized social group in a positive context may reduce prejudice toward that group over time (e.g., **Binder, Zagefka, Brown, Funke, Kessler, Mummendey, et al., 2009**) and may even reduce prejudice toward other out-groups in general (**Tausch, Hewstone, Kenworthy, Psaltis, Schmid, Popan, et al., 2010**). Reductions in implicit bias, specifically, have occurred as a result of longer-term exposure to minorities in socially valued roles (**Dasgupta & Rivera, 2008; Dasgupta & Asgari, 2004**), in the context of diversity education (**Rudman, Ashmore, & Gary, 2001**), and even as a result of simply imagining (rather than actually encountering) counter-stereotypes (**Blair, Ma, & Lenton, 2001**). In addition, some research indicates that people who have developed chronic egalitarian goals may be able to beat implicit bias at its own game by automatically inhibiting implicit stereotypes (e.g., **Moskowitz & Li, 2011**; **Moskowitz, Salomon, & Taylor, 2000**).

References

Banaji, M., & Heiphetz, L. (2010). Attitudes. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology, 5th edition* (pp. 348-388). New York: Wiley.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman.

Baron, A., & Banaji, M. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*, 53-58.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*, 991-1013.

Binder, J., Zagefka, H., Brown, R., Funke, F., Kessler, T., Mummendey, A., et al. (2009). Does contact reduce prejudice or does prejudice reduce contact? A longitudinal test of the contact hypothesis among majority and minority groups in three European countries. *Journal of Personality and Social Psychology, 96, 843*-856.

Blair, I. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.

Blair, I., Ma, J., & Lenton, A. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828-841.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27-41.

Blanton, H., & Jaccard, J. (2008). Unconscious racism: A concept in pursuit of a measure. *Annual Review of Sociology, 34*, 277-297.

Blascovich, J., & Mendes, W. (2010). Social psychophysiology and embodiment. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology, 5th edition*. New York: Wiley.

Brendl, C., Chattopadhyay, A., Pelham, B., & Carvallo, M. (2005). Name letter branding: Valence transfers when product specific needs are active. *Journal of Consumer Research, 32*, 405-415.

Casey, P., Warren, R., Cheesman, F., & Elek, J. (2012). *Helping courts address implicit bias: Resources for education*. Williamsburg, VA: National Center for State Courts.

Correll, J., Park, B., Judd, C., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*, 1314-1329.

Correll, J., Park, B., Judd, C., Wittenbrink, B., Sadler, M., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92*, 6, 1006-1023. doi:10.1037/0022-3514.92.6.1006

Cunningham, W., Johnson, M., Gatenby, J., Gore, J., & Banaji, M. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology, 85*, 639-649.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40*, 642-658.

Dasgupta, N., DeSteno, D., Williams, L., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice. *Emotion, 9*, 585-591.

Dasgupta, N., Greenwald, A., & Banaji, M. (2003). The first ontological challenge to the IAT: Attitude or mere familiarity? *Psychological Inquiry, 14*, 238-243.

Dasgupta, N., & Rivera, L. (2008). When social context matters: The influence of long-term contact and short-term exposure to admired outgroup members on implicit attitudes and behavioral intentions. *Social Cognition, 26,* 112-123.

De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology, 50*, 77-85.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 680-690.

Dovidio, J., & Gaertner, S. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11*, 315-319.

Dovidio, J., Gaertner, S., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology, 8,* 88-102.

Eberhardt, J., Davies, P., Purdie-Vaughns, V. & Johnson, S. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*, 383-386.

Eberhardt, J., Goff, P., Purdie, V., & Davies, P. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*, 876-893.

Fazio, R., Jackson, J., Dunton, B., & Williams, C. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.

Fazio, R., & Olson, M. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Fazio, R., Sanbonmatsu, D., Powell, M., & Kardes, F (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229-238.

Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science, 321*, 1100-1102.

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology, 50,* 141-150.

Gazal-Ayal, O., & Sulitzeanu-Kenan, R. (2010). Let my people go: Ethnic in-group bias in judicial decisions - Evidence from a randomized natural experiment. *Journal of Empirical Legal Studies, 7*, 403-428.

Gilbert, D., & Hixon, J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509-517.

Green, A., Carney, D., Pallin, D., Ngo, L., Raymond, K., Iezzoni, L., & Banaji, M. (2007). The presence of implicit bias in physicians and its prediction of thrombolysis decisions for Black and White patients. *Journal of General Internal Medicine, 22*, 1231-1238.

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

Greenwald, A., Banaji, M., Rudman, L., Farnham, S. Nosek, B., & Mellot, D. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3-25.

Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Greenwald, A., Nosek, B., & Banaji, M. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

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

Greenwald, A., Smith, C., Sriram, N., Bar-Anan, Y., Nosek, B. (2009). Implicit race attitudes predicted vote in the 2008 U.S. presidential election. *Analyses of Social Issues and Public Policy, 9*, 241-253.

Ito, T., & Cacioppo, J. (2007). Attitudes as mental and neural states of readiness. In B. Wittenbrink and N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 125-158). New York: Guilford.

Ito, T., Thompson, E., & Cacioppo, J. (2004). Tracking the timecourse of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin, 30*, 1267-1280.

Jones, J., Pelham, B., Carvallo, M., & Mirenberg, M. (2004). How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology, 87*, 665-683.

Jost, J., Rudman, L., Blair, I., Carney, D., Dasgupta, N., Glaser, J., et al. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior, 29,* 39-69.

Joy-Gaba, J., & Nosek, B. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology, 41*, 137-146.

Karpinski, A., & Hilton, J. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774-788.

Kitayama, S., & Uchida, Y. (2003). Explicit self-criticism and implicit self-regard: Evaluating self and friend in two cultures. *Journal of Experimental Social Psychology, 39*, 476-482.

Lemm, K., Sattler, D., Khan, S., Mitchell, R., & Dahl, J. (2002). *Reliability and validity of a paper-based implicit association test*. Paper presented at the fourth annual convention of the Society for Personality and Social Psychology, Savannah, GA.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*, 981-993.

MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*,163–203.

Mahaffey, A., Bryan, A., & Hutchison, K. (2005). Using startle eye blink to measure the affective component of antigay bias. *Basic and Applied Social Psychology, 35*, 537-545.

Mitchell, G., & Tetlock, P. (2006). Antidiscrimination law and the perils of mindreading. *Ohio State Law Journal, 67*, 1023-1122.

Mitchell, G., & Tetlock, P. (2009). Facts do matter: A reply to Bagenstos. *Hofstra Law Review, 37*, 737-762.

Moskowitz, G., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology, 47*, 103-116.

Moskowitz, G., Salomon, A., & Taylor, C. (2000). Preconsciously controlling stereotyping: Implicitly activated egalitarian goals prevent the activation of stereotypes. *Social Cognition, 18*, 151-177.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264–336).Hillsdale, NJ: Lawrence Erlbaum Associates.

Nosek, B. (2007). Implicit-explicit relations. *Current Directions in Psychological Science, 16*, 65-69.

Nosek, B., & Banaji, M. (2001). The Go/No-go Association Task. *Social Cognition, 19*, 626-666.

Nosek, B., Greenwald, A., & Banaji, M. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

Olson, M., & Fazio, R. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413-417.

Olson, M., & Fazio, R. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*, 636-639.

Payne, K., Cheng, C., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293.

Payne, K., Krosnick, J., Pasek, J., Leikes, Y., Ahktar, O., & Thompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Psychology, 46*, 367-374.

Pelham, B., Mirenberg, M., & Jones, J. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology, 82*, 469-487.

Penner, L., Dovidio, J., West, T., Gaertner, S., Albrecht, T., Dailey, R., & Markova, T. (2010). Aversive racism and medical interactions with Black patients: A field study. *Journal of Experimental Social Psychology, 46*, 436-440.

Phelps, E., O'Connor, K., Cunningham, W., Funayama, E., Gatenby, J., Gore, J., & Banaji, M. (2000). Performance on indirect measures of race evaluation predicts amygdale activation. *Journal of Cognitive Neuroscience, 12*, 729-738.

Plant, E., & Peruche, B. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science, 16*, 180-183.

Plant, E., Peruche, B., & Butz, D. (2005). Eliminating automatic racial bias: Making race non-diagnostic for responses to criminal suspects. *Journal of Experimental Social Psychology, 41*, 141-156.

Rankin, R., & Campbell, D. (1955). Galvanic skin response to Negro and white experimenters. *Journal of Abnormal and Social Psychology, 51*, 30-33.

Rooth, D. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics, 17*, 523-534.

Rudman, L. (2004). Social justice in our minds, homes, and society: The nature, causes, and consequences of implicit bias. *Social Justice Research, 17*, 129-142.

Rudman, L., Ashmore, R., & Gary, M. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81*, 856-868.

Schwarz, N., & Clore, G. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513-523.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology, 39*, 75-82.

Shields, S., & Harriman, R. (1984). Fear of male homosexuality: Cardiac responses of low and high homonegative males. *Journal of Homosexuality, 10*, 53-67.

Sinclair, S., Dunn, E., & Lowery, B. (2005). The relationship between parental racial attitudes and children's implicit prejudice. *Journal of Experimental Social Psychology, 41*, 283-289.

Sommers, S., & Ellsworth, P. (2001). White juror bias: An investigation of prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law, 7*, 201-229.

Son Hing, L., Li, W., & Zanna, M. (2002). Inducing hypocrisy to reduce prejudicial responses among aversive racists. *Journal of Experimental Social Psychology, 38*, 71-78.

Stanley, D., Phelps, E., & Banaji, M. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science, 17*, 164-170.

Stelfox, H., Gandhi, T., Orav, E., & Gustafson, M. (2005). The relation of patient satisfaction with complaints against physicians and malpractice lawsuits. *American Journal of Medicine, 118*, 1126-1133.

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Tausch, N., Hewstone, M., Kenworthy, J., Psaltis, C., Schmid, K., Popan, J., et al. (2010). Secondary transfer effects of intergroup contact: Alternative accounts and underlying processes. *Journal of Personality and Social Psychology, 99*, 282-302.

Vanman, E., Saltz, J., Nathan, L., & Warren, J. (2004). Racial discrimination by low-prejudiced Whites: Facial movements as implicit measures of attitudes related to behavior. *Psychological Science, 15*, 711-714.

Vargas, P., Sekaquaptewa, D., & von Hippel, W. (2007). Armed only with paper and pencil: "Low-tech" measures of implicit bias. In B. Wittenbrink and N. Schwarz (Eds.), *Implicit measures of attitudes*, pp. 103-124. New York: Guilford.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology, 33*, 490-509.

Wax, A., & Tetlock, P. (2005, December 1). We are all racists at heart. *Wall Street Journal*, A16.

Wittenbrink, B., & Schwarz, N., Eds. (2007). *Implicit measures of attitudes*. New York: Guilford.

Ziegert, J., & Hanges, P. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology,* 90, 553-562.