

Policy Implications of Implicit Social Cognition

Brian A. Nosek

Rachel G. Riskind

University of Virginia

Contact information

Brian Nosek
Box 400400; 102 Gilmer Hall
Department of Psychology
University of Virginia
Charlottesville, VA 22904-4400
nosek@virginia.edu

Authors' Note: This project was supported by Project Implicit, Inc. Nosek is an officer of Project Implicit, a non-profit organization that includes in its mission “To develop and deliver methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race, gender or other factors.”

Abstract

Basic research in implicit social cognition demonstrates that thoughts and feelings outside of conscious awareness or conscious control can influence perception, judgment and action. Implicit measures reveal that people possess implicit attitudes and stereotypes about social groups that are often distinct from their explicitly endorsed beliefs and values. The evidence that behavior can be influenced by implicit social cognition contrasts with social policies that implicitly or explicitly assume that people know and control the causes of their behavior. We consider the present state of evidence for implicit social cognition and its implications for social policy. We conclude that considering implicit social cognition can contribute usefully to policy, but that most uses of implicit measures themselves as selection or evaluation devices is not easily justified.

Keywords: implicit social cognition, Implicit Association Test, validity, public policy, stereotypes, discrimination

Two revelations from the mind sciences are easy to understand about other people's minds yet difficult to accept about one's own: (1) much of mental life occurs outside of conscious awareness and conscious control, and (2) thinking is often not objective or rational. Minds use shortcuts, heuristics, inferences, stereotypes, and expectations to help comprehend what happened before and predict what will happen next. These processes can be useful and effective on one occasion, and suboptimal or plain wrong on another. Moreover, they can influence social perception, judgment and behavior implicitly – without the person's awareness, intention or control.

If this implicit social cognition were merely a quirky feature of the mind's internal dialogue prior to conscious intentions forming perceptions, making judgments, or initiating actions, then it would not have much relevance for social policy. Social policies are principles, guidelines, and legislation that affect human welfare. Such policies rarely address how humans think; instead, they address how humans behave. Of course, thinking and behaving are not independent; implicit (and explicit) cognition shapes perception, judgment, and action whether the actor is aware of the influence or not. As such, social policies may be ineffective or counterproductive when they assume (a) that human behavior is necessarily a function of conscious intent, and (b) that people are aware of the causes of their behavior. Effective design, implementation and evaluation of social policy require realistic models of the mind in order to best fulfill their purposes.

Research in implicit social cognition can impact social policy in at least two ways. First, practitioners who design and evaluate policy could apply theory and evidence of determinants of social behavior that subvert conscious intention, awareness, or control. Second, practitioners might apply measurement methods of implicit social cognition themselves, such as the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), evaluative priming (Fazio et al., 1986), and their derivatives, as part of policy applications.

We begin this article by outlining our perspective on the role of basic science in policy design and evaluation. Next, we review the accumulated evidence about implicit social cognition with relevance for social policy. Finally, we consider opportunities and cautions regarding applications of this evidence to social policy.

The policy implications of basic science

As a discipline of psychology, implicit social cognition is a basic science enterprise. That is, the research is defined by the fundamental knowledge it develops without direct consideration of practical implications. However, as often occurs, basic research about implicit social cognition has become the basis of applied science for a variety of topics (e.g., Lee et al., 2009; Molesworth & Chang, 2009; Petroczi et al., 2010; Teubel et al., 2011), and for social application in law (Greenwald & Krieger, 2006; Kang & Banaji, 2006; Kang & Lane, 2009; Rachlinski et al., 2009), organizational practices (Agerstrom & Rooth, 2011; Chugh, 2004; Leavitt et al., 2011; Rooth, 2010; Srivastava & Banaji, 2011), education (Nosek et al., 2009; Cvencek et al., 2010), and health (Green et al., 2007; Schwartz et al., 2006; Teachman et al., 2010; Wiers et al., 2010). Potential connections between the basic and applied implicit social cognition are intuitive, resulting in a rapid accumulation of debate and discussion about how to apply research in implicit social cognition (Bagenstos, 2007; Bartlett, 2009; Greenwald, 2008; Jolls & Sunstein, 2006; Jost et al., 2009; Kang, 2006; Krieger & Fiske, 2006; Landy, 2008; Mitchell & Tetlock, 2006; Rudman, 2008; Tetlock and Mitchell, 2009).

Basic researchers often adopt a defensive crouch on the application of their research to policy. Common protestations include “we do not yet know enough to say anything about application” and “there is too much uncertainty to say anything useful” (Gardner & Scherer, 1990). Such concerns may lead basic researchers to avoid saying anything at all.

This is unfortunate. Policy decisions are being made today. If the basic scientist, the person most expert on the problem, is unwilling to contribute knowledge or critique proposed applications, then what is the basis of the implementation? In the absence of scientific evidence, practitioners’ intuitions, folk theories, ideologies, and assumptions will fill the void to determine policy. We argue that the basic scientist’s responsibility is to represent the existing state of knowledge as accurately and informatively as possible: uncertainties, cautions, and all. The scientist should *not* decide when scientific evidence is ready to be shared or declare a moratorium on application until the problem is solved. After all, how many scientific papers end with the phrase “no more research is necessary”? (The answer is zero.)¹ Policy formulation does not demand the absolute truth; it demands the *best available* evidence (Goodman, Levine, & Melton, 1992). As Bazelon (1982) admonished, “Do not think...that the system will be any more humane in your absence.”

Simultaneously, the basic scientist must appreciate her own limits of knowledge and avoid overstepping. The scientist is responsible for representing the state of knowledge; the practitioner is responsible for considering the knowledge, its uncertainties, and the context of application to make the best possible policy decisions (Bazelon, 1982; Schauer, 2010). Because basic scientists are accustomed to experimental control and isolating targeted variables, they sometimes fail to appreciate the *context* of policy applications. Basic scientists may also overestimate the potency of their favorite variables in the real world, in which there are many moderating and intervening influences that were not varied in the experimental context. Application of basic science to policy is messy, making it quite reasonable for the fastidious basic scientist to worry about contamination. However, without scientific theory and evidence, policy formulation will still proceed. Life does not wait for the basic scientist; real social problems are immediate and urgent. Bazelon (1982) offered cogent wisdom about the scientist’s contribution to policy:

“The public arena ... is a place where awesome decisions must be made on the basis of limited knowledge and groping understanding. The mindset I urge you to bring to the public arena should therefore combine the scientist’s ease in admitting uncertainty and the clinician’s tolerance of pragmatic decision-making. It should combine the scientist’s rigor with the clinician’s appreciation of the broader human context. I know that this is a tall order, but I also know that nothing else is acceptable” (p. 119-120).

Generalization in science and to policy

Empirical reports do not include statements like “These findings were observed under these particular conditions with this particular sample in this setting at this point in history, but they will never be observed again and have no implications beyond characterizing that moment in time for its own sake.” All empirical research is generalized beyond the original data and findings. When making generalizations, scientists ask: “to what extent does the evidence

¹ The phrase does appear in a few articles – almost entirely to point out that it never appears in articles, or that it is not the basis for deciding when science is ready to be applied (e.g., Fulero, 1986; Kitchell et al., 2006).

generated from a given demonstration apply to anything other than the demonstration itself?” This question is critically important for the bridge between scientific evidence and policy applications.

One aspect of generalization addresses the conditions under which the findings from one study can be replicated in another study. When a study demonstrates a statistically significant result, researchers usually assume that the same result would recur if a second study replicated critical elements of that investigation. The generalizability challenge is to identify the “critical elements”.²

Imagine, for example, a study measuring how far participants choose to sit from another person. The researchers randomly manipulate the person to be either thin or obese and find a reliable difference – people sit farther away from an obese person than a thin person (similar to Bessenoff & Sherman, 2000). Which features of the study cause the finding? Would the same effect be observed with blue or pink walls, in January as in May, if the confederates wore pants instead of shorts, or if the chairs were stools instead? It would be quite surprising if the finding did not generalize across such variations. However, that does not mean that it can be generalized across all circumstances.

It is impossible to test more than a trivial number of the infinite variety of social circumstances. As such, qualitative considerations like logic, reasoned argument, and existing knowledge play a substantial role in identifying the boundaries of acceptable generalization. Would the same effect be observed (a) with participants who differed from original participants in age, education, or weight status; (b) if social proximity were assessed by willingness to touch, as opposed to seating distance; (c) after establishing a particular mindset in participants, such as screening Martin Luther King Jr’s “I Have a Dream” speech immediately prior to the key measurement? In these cases, common intuitions suggest that variation could exert important influences on the original effect. But, because social circumstances vary infinitely, generalization is the default presumption. Qualitative considerations identify plausible constraints or moderating influences, and scientific research progresses by evaluating the plausible moderating influences.

This discussion is situated in the conducting of scientific research, but the same considerations apply to the translation of scientific findings to policy application. Are there circumstantial factors in where, when, and how the policy is, or would be, practiced that affect the applicability of the relevant scientific evidence? A common trope for criticizing the application of scientific evidence produces a resounding “yes” to this question. Most scientific research does not look anything like real-life policy. Scientific research settings are artificial; the measures are different than the behaviors that occur in practice; and the participant samples are not the ones who are directly affected by the policy practices. It is rare to find research on the policies as practiced, in the settings that they are practiced, and with the people that are practicing them. And, even when it is done, a committed skeptic can still identify many differences between the research findings and the exact policy circumstances under consideration.

In most cases, however, the question of generalization does not concern the research design, procedure, and measures themselves. The question is whether the psychological processes that are identified are likely to be operating in the other settings as well. Mook (1983) provides illustration of this point by considering a visit to the optometrist. It is patently obvious

² Another issue for replicability is the possibility that the original finding was a false positive; the test of statistical significance suggested that the effect did not occur by chance.

that the procedures and practices during an eye appointment do not occur in real life. The goal is not to see how well you will identify light flashes looking into a mini dome with an eye patch over one eye when you confront that same situation in daily life. Instead, the procedures are designed to isolate and clarify the health and operation of the visual system. It is not the eye procedures themselves that are generalized; it is the findings and processes that are revealed by those procedures.

Reasonable people can disagree on the plausibility of particular threats to generalization, but accumulation of evidence can make some factors much less plausible than others. For policy application, the scientist's job is to outline the relevant evidence, identify the status of uncertainty in the findings, and clarify – as best as the evidence can suggest – the opportunities and threats to generalizability between the basic results and the proposed application. With this in mind, we now turn to defining implicit social cognition, summarizing some of the cumulative research evidence that is relevant for policy application, and identifying opportunities for the basic science to inform policy.

What is implicit social cognition?

Implicit social cognition is a descriptive term encompassing thoughts and feelings that occur independently of conscious intention, awareness or control (Greenwald & Banaji, 1995; Nosek & Greenwald, 2009; Nosek, Hawkins, & Frazier, 2011). This term does not refer to a specific psychological process. Rather, implicit social cognition is an inclusive term that is a direct contrast from explicit social cognition: thoughts and feelings that are a product of intention, are consciously available, are controlled, and require mental resources (Bargh, 1994; see Moors & De Houwer, 2006 for an alternative taxonomy of features of automaticity and implicit cognition). The absence of any of these features suggests that the behavior is not fully explicit. A consequence of the descriptive definition is that implicit social cognition is a heterogeneous concept that is inclusive of distinct mechanisms.

This offers both benefits and challenges for considering policy implications of implicit social cognition. For some policy matters, there may be different implications – particularly normative implications – depending on the specific contributing psychological processes. For example, would culpability differ for a person who was aware of a psychological process producing a behavior (but could not control this process), versus a person who was unaware that the process was operating in the first place? If this difference in psychological states is material for policy implications, then distinction between psychological processes must be taken into account. If, however, psychological states are immaterial and policy is relevant only to the ultimate behavior, then molar view of how implicit social cognition influences behavior may be sufficient, even preferable by not committing to particular psychological processes.³

Implicit social cognition research emphasizes the assessment of social cognition with procedures that do not require awareness of the association between the attribute of interest and the response generated (Nosek & Greenwald, 2009). Thoughts and feelings are assessed without

³ Relatedly, it is common to have a relatively naïve understanding of the operative processes of measurement but still have a very productive understanding of its usefulness in predicting behavior. For example, the processes underlying conscious introspection for answering survey questions are largely a mystery, but survey measurement is used productively for predicting and understanding behavior (Nosek & Greenwald, 2009).

the respondent deliberately introspecting and generating a report. As a consequence, implicit measures can reveal mental content that respondents are otherwise unwilling or unable to report.

Consider, for example, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; Nosek, Greenwald, & Banaji, 2007), the most popular measure in implicit social cognition research (Nosek et al., 2011). In the IAT, items representing one of four categories such as *male*, *female*, *career*, and *family* appear on a computer screen one at a time. Participants must categorize the items into their respective category using one of two keys on the keyboard as quickly as they can. There are two critical conditions of the task. In one condition, participants categorize items representing *male* (e.g., he, boy) and *career* (e.g., business, salary) with one key, and items representing *female* (e.g., she, girl) and *family* (e.g., children, home) with the other key. In the other condition, participants categorize items representing *female* and *career* with one key, and *male* and *family* with the other key. The implicit measure is a comparison of the average response time categorizing items between the two conditions. The logic of the task is straightforward; people should be faster to categorize the items when the categories sharing a response key are associated mentally than if they are not associated. In this case, most people are faster categorizing male and career (and female and family) with the same key compared to the reverse (Nosek, Smyth, et al., 2007). This is interpreted as an implicit stereotype because participants do not consider whether they differentially associate the genders with career and family; they just categorize the items as quickly as they can. Respondents have no opportunity to consider and endorse or reject the thoughts that influence performance on the IAT. Implicit measures, like the IAT, can reveal associations that the respondent would genuinely reject as invalid if the opportunity were available.

Claims about implicit social cognition with a strong foundation of evidence

The history of implicit social cognition includes both theoretical developments about non-conscious or automatic thoughts and empirical demonstrations of automaticity and mental operations that occur outside of awareness (e.g., Bargh, 1994; Greenwald, 1990; Freud, 1900; Helmholtz, 1910/1925; Nisbett & Wilson, 1977; Schneider & Schiffrin, 1977; Schiffrin & Schneider, 1977). Investigation of implicit social cognition increased dramatically during the last 20 years with the emergence of effective implicit measurement methods (e.g., Fazio et al., 1986; Greenwald et al., 1998; Nosek & Banaji, 2001; Payne et al., 2005; for reviews of implicit measures see De Houwer et al., 2009; Fazio & Olson, 2003; Nosek, et al., 2007; Nosek, et al., 2011; Teige-Mociemba et al., 2010; Wentura & Degner, 2010). Here, we consolidate the considerable research literature to what, we believe, are claims that have a strong basis of empirical support about implicit social cognition (for more detailed reviews of implicit social cognition see Gawronski & Payne, 2010; Greenwald & Banaji, 1995; Nosek, Hawkins, & Frazier, in press).

It is now received wisdom that (*Claim 1*) *much of mental life occurs outside of conscious awareness or control*. Processes that occur independently of conscious experience and conscious intention shape social perception, judgment, and action (Bargh, 1994; Bargh & Chartrand, 1999; Nisbett & Wilson, 1977; Wegner, 2002; Wilson, 2002). Moreover, people lack complete control of what enters their minds. (*Claim 2*) *Exposure to information, even exposure brief enough to avoid being consciously experienced, is sufficient for information to be encoded and influential* (Draine & Greenwald, 1999; Kunst-Wilson & Zajonc, 1980; Murphy & Zajonc, 1993; Zajonc, 1968). Zajonc (1968) flashed images of Chinese pictographs so briefly that participants did not detect their appearance. Later, when asked to decide which of a set of

pictographs they preferred, participants were more likely to select the ones that had been presented earlier. Subliminal exposure to this information was sufficient to alter participants' evaluations without their knowledge of the source or its influence.

As a consequence, (*Claim 3*) *people may possess implicit associations with which they actively and honestly disagree* (Banaji, 2001; Banaji et al., 2004; Devine, 1989; Gawronski & Bodenhausen, 2006). For example, many people who report egalitarian values and attitudes possess implicit associations linking the concept *good* with *young people*, *White people*, *thin people*, and *straight people* more easily than with *old people*, *Black people*, *fat people*, and *gay people*, respectively (Cunningham, Nezlek, & Banaji, 2004; Nosek et al., 2002; Nosek, Smyth, et al., 2007). These may form in memory via mechanisms that are distinct from the formation of explicit attitudes and beliefs (Rydell & McConnell, 2006; Ranganath & Nosek, 2008). In particular, implicit associations may be sensitive to contingencies observed in the environment, regardless of whether one is aware of, or agrees with, those contingencies (Gawronski & Bodenhausen, 2006).

The most popular topic of investigation in implicit social cognition research is the evaluation of social groups. This research reveals that (*Claim 4*) *implicit preferences for some social groups over others are pervasive* across hundreds of studies and millions of participants (Nosek, Banaji, & Greenwald, 2002; Nosek, Smyth, et al., 2007).⁴ Nosek and colleagues (2007) summarized findings from more than a dozen social evaluations with more than four million completed implicit measures in total and observed that implicit social preferences were present across age, gender, race/ethnicity, education, region, and ideology. Notably, however, while evidence has accumulated with millions of participants, few studies have employed representative samples in order to obtain population estimates of implicit associations.

Across the accumulated data, (*Claim 5*) *the magnitude of implicit preferences varies by topic, person, and situation*. Some topics elicit strong average implicit preferences, such as evaluations of young compared to old people, and others elicit much weaker *average* implicit preferences, such as evaluations of Democrats compared to Republicans (Lindner & Nosek, 2009; Nosek et al., 2002). Simultaneously, implicit preferences vary substantially across individuals for any given topic. For example, Democrats tend to implicitly prefer Democrats, to varying degrees, and Republicans tend to implicitly prefer Republicans (Lindner & Nosek, 2009; Nosek et al., 2002; Nosek, Banaji, & Jost, 2009). Further, different implicit evaluations for the same content may be elicited depending on the social situation or context (Blair, 2002), a topic that will be taken up in more detail in the next section.

(*Claim 6*) *Implicit social cognition is distinct, but related to self-reported evaluation of the same content* (Cunningham et al., 2001; Greenwald & Farnham, 2000; Nosek & Smyth, 2007). Correlations of individual differences between the two assessments range from weakly to strongly positive across topics (Nosek, 2007). For example, implicit and explicit measures of preferences for young, as opposed to old, people are weakly correlated ($r \sim .10$), while implicit and explicit measures of preferences for Democrats, as opposed to Republicans, are strongly correlated ($r \sim .60$).

⁴ Not everyone agrees. Blanton and colleagues (2006) speculate that IAT measurement does not have a rational zero-point and estimates of implicit preferences for majority or socially-favored groups may be exaggerated. However, others argue that the available empirical examinations have supported interpretations of the IAT zero-point as indicating no difference in association strengths between the two response conditions (Greenwald, Sriram, & Nosek, 2006; Greenwald et al., 2002; Greenwald & Nosek, 2009).

Implicit and explicit measures might elicit different effects if the former reflects cultural knowledge, whereas the latter reflects one's own evaluations. However, (*Claim 7*) *variation in correspondence between implicit and explicit (self-report) assessments is not attributable to the former being the culture's evaluation and the latter being one's personal evaluation* (Gawronski et al., 2008; Nosek & Hansen, 2008; Uhlmann, Poehlman, & Nosek, 2011).⁵ There is substantial evidence, however, that (*Claim 8*) *variation in implicit-explicit correspondence across topics and individuals is predicted by features of the evaluation*. For example, when self-presentation concerns are higher, such as for attitudes toward racial/ethnic groups as compared to political groups, implicit-explicit correspondence tends to be weaker (Fazio et al., 1995; Nosek, 2005). Moreover, stronger or more important attitudes elicit stronger implicit-explicit correspondence than weaker attitudes (Karpinski, Steinman, & Hilton, 2005; Nosek, 2005).

Ultimately, the ways that implicit social cognition influences behavior are key for informing policy. For one, (*9*) *people do not observe their mental operations*, making the introspective assessment of the reasons for one's behavior interesting, but indeterminately related to actual causes of the behavior. People can be simultaneously confident and wrong about the reasons for their behavior (Bem, 1972; Nisbett & Wilson, 1977). Also, as implied already, (*10*) *conscious intention, and automatic or unconscious processes both influence behavior* (Bargh, 1994; Bargh & Chartrand, 1999; Chaiken & Trope, 1999). In some cases, self-reported evaluations predict better than implicit measures; in other cases, implicit measures predict better, particularly in socially sensitive circumstances (Greenwald et al., 2009). Finally, (*11*) *both implicit and explicit social cognition predict variation in human behavior that is not predicted by the other*.⁶ Implicit and explicit measures are not redundant; each can contribute independently for predicting human behavior (Greenwald et al., 2009). In total, these claims provide a foundation of evidence for implicit social cognition's construct and predictive validity.

Uncertainties in implicit social cognition

Science is cumulative. New evidence builds on, clarifies, or modifies prior evidence, leading to a more accurate accounting of reality. This is a process of uncertainty reduction. Over time, models improve at predicting and explaining behavior, and the portion that is not explained diminishes. There may never be a perfect model of behavior.⁷ Even so, a goal of

⁵ While we believe the empirical evidence for this claim is strong, not everyone agrees (Arkes & Tetlock, 2004; Han, Olson & Fazio, 2006; Karpinski & Hilton, 2001; Olson & Fazio, 2004). Further, there is an interesting debate on whether such cultural knowledge would be considered an extraneous influence on implicit measurement or part of implicit social cognition (Olson & Fazio, 2004; Gawronski et al., 2008; Nosek & Hansen, 2008).

⁶ Not everyone agrees. Blanton and colleagues (2009) suggest that the predictive validity evidence for the IAT, particularly with regard to racial attitudes, is specious. They reanalyzed data from two published studies to support their claims (McConnell & Leibold, 2001; Zeigart & Hanges, 2005). The original authors disagreed with Blanton and colleagues analytical criticisms (McConnell & Leibold, 2009; Zeigart & Hanges, 2009).

⁷ Isaac Asimov's Foundation series imagined a future in which psychological models became so advanced that behavior on a large scale could be predicted in advance. As one might predict, the models still failed in vitally important ways.

science is to generate the simplest possible models, making them comprehensible, that account for the most variation in behavior, making them powerful.

The prior section delineated eleven claims that, in our view, have a firm foundation of evidence and relatively certain validity. This section describes research evidence about implicit social cognition that either (a) highlights uncertainties in measurement and prediction or (b) reflects the need for continued elaboration of theory and evidence. Understanding such areas of scientific uncertainty will help clarify the appropriate application of implicit social cognition to policy.

Implicit measures are influenced by extraneous factors. No psychological measure is a pure assessment of its intended construct. There is always uncertainty in measurement, due to imperfect internal consistency and extraneous influences. One source of uncertainty for implicit measures is that assessment is usually indirect; the construct of interest is inferred based on performance of another task. In the Affect Misattribution Procedure (AMP; Payne et al., 2005), for example, participants rate a series of unfamiliar Chinese ideographs as pleasant or unpleasant. Immediately prior to presenting the ideographs, a prime image or word appears briefly. Despite instructions to ignore them, primes systematically influence ratings of the ideographs. The implicit measure assesses the indirect influence of the prime on evaluation of the ideographs. However, to the extent that the ideographs being judged have any influence on the ultimate evaluation, the measure is not a “pure” assessment of evaluations of the primes. As a consequence, indirect measures tend to have lower internal consistency than direct measures. Implicit measures have internal consistencies ranging from very low, as in sequential priming procedures (Cunningham et al., 2001; Olson & Fazio, 2003), to moderate or strong with the AMP, IAT, and variants of the IAT (Nosek et al., 2007; Nosek et al., in press; Payne et al., 2005).

The IAT and other procedures that compare average response latencies between two or more conditions are also influenced by extraneous procedural factors. For example, slower average response latencies are associated with larger differences between response conditions (Cai, Sriram, Greenwald, & McFarland, 2004; McFarland & Crouch, 2002; Mierke & Klauer, 2003; Sriram, Greenwald, & Nosek, 2009). Also, the critical response condition performed first tends to interfere with the one performed second (Greenwald et al., 1998).

Extraneous influences may reduce reliability and certainly have a negative impact on construct and predictive validity. As such, identification of extraneous influences is an important step toward reducing their impact (see Nosek, Greenwald, & Banaji, 2007 for a review of extraneous influences on the IAT in particular). For example, a procedural adaptation for the IAT can reduce the extraneous influence of pairing order (Nosek et al., 2005), and an analytic improvement can reduce the extraneous influence of overall average latency (Greenwald et al., 2003; Mierke & Klauer, 2003). Further, analytic practices may allow researchers to isolate and nullify extraneous influences (Cunningham, Nezlek, & Banaji, 2004; Nosek & Smyth, 2007). Relatedly, multiple processes contribute to performance on implicit measures (Conrey et al., 2005), and analytic or procedural innovations may isolate the key components of task performance (Sherman, Klauer & Allen, 2010). To the extent that extraneous factors influence assessment, there will be greater uncertainty in interpretation of the effect.

Implicit social cognition is malleable. One way to think about a person’s implicit attitudes, identities, and stereotypes is to imagine that they are akin to personality. The popular intuition is that personality is relatively fixed across time and circumstance. This intuition is incorrect (Mischel, 1973; Mischel & Shoda, 1995). Personality is contextualized. A person may

be relatively more introverted in some circumstances than others, and unique combinations of social settings may elicit unique behavioral responses.

Implicit social cognition is likewise contextualized (Blair, 2002; Gawronski & Sritharan, 2010). Despite their automatic generation, implicit evaluations are sensitive to social circumstances, such as contextual clues about the valence of the social target (Barden et al., 2004), viewing positive or negative instances prior to evaluating a social category (Dasgupta & Greenwald, 2001; Joy-Gaba & Nosek, 2010), participants' motivation to affiliate with another person whose beliefs are readily apparent (Sinclair, Lowery, Hardin, & Colangelo, 2005), or even mental and physiological states such as implicit evaluations of smoking among habitual smokers following relative deprivation from smoking (Sherman et al., 2003; Payne et al., 2007).

The evidence for malleability has produced unresolved (perhaps unresolvable) debates about whether implicit social cognitions are better thought of as things that exist or outcomes that are constructed (Fazio, 1990; Martin & Tesser, 1992; Mitchell, Nosek & Banaji, 2003; Smith & DeCoster, 2000; Schwarz, 1999; Schwarz & Bohner, 2001). Regardless of one's theoretical positioning in that debate, it is clear that implicit measurement elicits some degree of stability over time (trait-like) and some degree of malleability across time and circumstance (state-like). For example, the average test-retest reliability for the IAT across studies was .56, an estimate that varied little as a function of the time between the two tests (Nosek et al., 2007). The key implication for policy considerations is that implicit social cognition is not a fixed entity; it is dynamically interrelated with the social situation. The social circumstances may have implications for what is measured and how implicit social cognition predicts behavior.

How and when does implicit social cognition predict behavior? Accumulated evidence leaves no doubt that implicit measures predict behavior (Greenwald et al., 2009), but this evidence has not yet converged into a model accounting for *how* and *when* implicit social cognition predicts behavior. Nosek, Hawkins, and Frazier (2011; in press) reviewed the existing theories and evidence and suggested that implicit social cognition is more likely to predict behavior when the individual lacks (a) motivation to deliberately direct the behavior; (b) opportunity to initiate the relevant behavior; (c) ability to control the behavior; or (d) awareness of factors that influence the behavior.

The context of interracial interaction illustrates these factors. People may possess motivation, opportunity, ability, and awareness to control what they say in an interaction with a person from another racial group, but they may lack motivation, opportunity, ability, or awareness to control their non-verbal behavior in that interaction. Dovidio, Kawakami, and Gaertner (2002) found that implicit, but not explicit, measures of racial attitudes predicted difficult-to-monitor and -control nonverbal behaviors, such as blinking rate and amount of eye contact (see also Fazio et al., 1995; McConnell & Leibold, 2001).

Simultaneously, some behaviors are ostensibly deliberate and controllable and can still be predicted by measures of implicit social cognition. For example, voting is perhaps the prototype of a controlled action: people must be motivated to initiate voting behavior, they have ample opportunity to decide their vote, and they can ensure that such a decision is embodied in their behavior. The prevailing assumption is that people are aware of the basis for their vote. Even so, implicit measures predict variation in voting behavior that self-report measures do not. For example, implicit racial preferences for white people compared to black people predicted reported vote in the 2008 U.S. Presidential election between Barack Obama and John McCain (Greenwald et al., 2009; Payne et al., 2010). Moreover, implicit measures of political positions predict undecided voters' ultimate votes (Galdi, et al., 2008; Rocco & Zogmaister, 2010). As

behavioral prediction is the area of most theoretical and practical interest, refinement of the conditions under which implicit measures predict behavior is likely to enjoy rapid progress in the next decade.

Of related interest is how strongly should implicit social cognition predict behavior for this association to be considered substantively interesting or important? In an interview with the *New York Times*, psychologist Philip Tetlock declared the candidacy (and ultimate election) of Barack Obama over John McCain in 2008 to be “a major embarrassment for the unconscious-bias crowd” because it should not have occurred if “unconscious bias were as pervasive and potent as typically implied” (Tierney, 2008). As the columnist asked, if more than 80% of whites have an implicit preference for Whites over Blacks, and “whites made up 74% of the voters . . . , why is Barack Obama going to be the next president?”

This implies that implicit racial bias ought to account for all of the variation in white people’s voting behavior. However, voting behavior, like most behavior, is influenced by multiple variables. Consider, for example, just a few of the relevant social biases that could have played a role in the choice between Obama and McCain – implicitly or explicitly. Age, height, and attractiveness biases would likely have favored Obama, while stereotypes associating the categories Americans and Presidents with White would have favored McCain. Of course, social biases are not the sole influences on voting behavior. The most obvious is ideology. A single-item self-rating of *strongly liberal* to *strongly conservative* accounts for 85% of the variation in presidential vote in elections between 1972 and 2004 (Jost, 2006). In other words, that single item leaves just 15% of the variation in voting behavior to be explained by other factors!

It is unlikely that any behavior is predicted perfectly by any single social bias, implicit or otherwise. So, how much *should* a social cognition predict a behavior for this association to be considered interesting or useful? This is a qualitative judgment and depends on the context of prediction: What is the behavior, what is the social cognition, what is the relevance of the social cognition to the behavior, and what other factors influence the behavior in that circumstance? In the case of voting for Barack Obama versus John McCain in 2008, the published empirical evidence with two different implicit measures (IAT and AMP) suggests that implicit racial attitudes account for small but statistically significant variations in vote, even after accounting for political ideology and self-reported racial attitudes (perhaps as little as 0.5% and as much as 3.5%; Greenwald et al., 2009; Payne et al., 2010). Is that “enough”? Ultimately, that decision is a value judgment. The descriptive fact is simply that implicit race bias accounted for variation in vote that was not accounted for by other measures.⁸

Variation across implicit measures. The nature and function of implicit social cognition is not defined by a single measurement method. A recent estimate suggests that almost half of the existing implicit social cognition literature used the IAT, another 20% used evaluative priming, and the remaining 30% is distributed across a variety of measures (Nosek, et al., 2011). Each measure’s idiosyncrasies constrain how the measured psychological constructs are understood. As such, diversity in measurement is important to identify what findings are idiosyncratic to measurement versus common across measures.

The literature is surprisingly sparse on comparative investigation of implicit measures. The first published study found very little commonality among implicit self-esteem measures (Bosson et al., 2000). Subsequent studies tend to observe stronger associations among implicit measures, especially after accounting for unreliability in measurement (e.g., Cunningham,

⁸ In both studies, self-reported racial attitudes also significantly contributed to predicting vote – as strongly or even more strongly than implicit attitudes did.

Preacher, & Banaji, 2001; Ranganath, Smith, & Nosek, 2008). Even so, it is clear that implicit measures may each reveal distinct aspects of implicit social cognition and may each carry irrelevant procedural idiosyncrasies. For policy, the implication is that there is not yet an account of the relations among implicit measures; each may contribute uniquely to assessing implicit social cognition. Findings observed with one may or may not be observed with another.

Effectiveness of interventions to change, prevent, control influence of implicit cognition on behavior. Possessing an implicit preference or stereotype does not guarantee that it will influence behavior. Even when conditions are aligned for a particular social cognition to influence judgment, people may engage corrective processes to avoid such an influence. Wilson and Brekke (1994) provide a model of debiasing that identifies the conditions necessary to eliminate unwanted thoughts from influencing judgment or behavior. To do so, people must (a) detect their unwanted thoughts, (b) be motivated to remove the biasing influence, (c) detect the size and direction of the biasing thoughts, and (d) redirect their own judgments or behaviors so that they are not contaminated by the biasing thoughts.

It is not yet clear to what extent implicit social cognition is amenable to prevention or alteration at each of these stages, but there is evidence that corrective processes can moderate behavior (see Bartlett, 2009). A variety of dual-process models articulate how implicit and explicit processes operate independently and interactively to produce judgment (Chaiken & Trope, 1999; Fazio, 1990; Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). The critical point, for the purposes of policy implications, is that having thoughts is not the same as acting on those thoughts. If corrective processes are effective, unwanted implicit social cognitions may be present in mind but fail to influence judgment and action. Based on the present literature, we are not particularly optimistic that explicit efforts to override or otherwise prevent the influence of unwanted implicit cognitions can be completely effective⁹, but the limits of debiasing through intention and explicit effort are still undetermined. This presents a ripe area for future research, and some uncertainty about the inevitability of implicit social cognition influencing behavior.

Applying implicit social cognition to social policy

The central implication of implicit social cognition is that social behavior is not just a product of *conscious* beliefs, values, and intentions. Thoughts and feelings that the individual may actively disagree with can exist in memory and influence behavior outside of awareness and control. How can policymakers apply this insight and the rest of theory and evidence about implicit social cognition?

At an abstract level, the principles of implicit social cognition can help evaluate whether policy and practice is behaviorally realistic (Kang & Banaji, 2006; Krieger & Fiske, 2006). In particular, people might assess a policy with the following questions. First, does the policy anticipate constraints of intention, awareness, and control on human behavior? If not, are there automatic processes that could subvert the goals of the policy? Second, can conditions ensure that actors have the awareness, opportunity, and ability to execute the desired course of action, if motivated to do so? If not, is there evidence that the likely automatic responses are in line with the policy objectives?

⁹ For example, people can believe that they have identified the causes of their behavior and/or engaged in corrective processes but, in fact, be wrong that they have done so (Nisbett & Wilson, 1977; Norton, Vandello, & Darley, 2004).

Third, for the policy to work, must the actors be aware of the conditions influencing their behavior? If so, how does the policy or practice ensure that people can identify these factors? Fourth, does the policy anticipate and account for contextual features that can influence behavior without the actors' recognition or control? If not, is it possible that such features could disrupt policy objectives? Finally, does evaluation of the policy's success rely on actors' self-assessments? For example, if a policy requires that the actor avoid using a target's gender in evaluation, is there any assessment of success other than the actor's report that he or she did not consider gender? Reviewing policy with these questions may elicit hypotheses or evidence that policy objectives are at risk because implementation and practices do not anticipate the role of implicit social cognition.

Next, we highlight several examples in which the influence of implicit social cognition in behavior is either implied or apparent. These applications are illustrative, not comprehensive, of opportunities to apply implicit social cognition.

Examples of implicit social cognition operating in law, health, organizational and civic behavior. The U.S. justice system employs a "rational actor" model of the mind. For example, potential jurors are asked if they can objectively review the facts and avoid biasing influences; this is a common criterion for inclusion on a jury. When inadmissible testimony is presented, a judge may instruct the jury to forget the information and not to allow it to influence their later judgment (Mueller & Kirkpatrick, 1995). Jurors may dutifully attempt to set this information aside, but research suggests that the mind is not always so accommodating (Kassin & Studebaker, 1998). In an experimental setting, the judgments of jurors who were instructed to put aside inadmissible information were closer to judgments from jurors who were allowed to use the information and differed most from judgments of jurors who never heard this inadmissible information in the first place (Kassin & Sommers, 1997).

Likewise, it is well understood that the social category and facial characteristics of criminal defendants are not supposed to influence judicial decisions, conviction or sentencing, but evidence suggests that they do (Griffin & Wooldredge, 2006; Johnson, 2006; Mustard, 2001). For example, an investigation of Arab and Jewish judges in Israel found evidence for bias favoring the in-group in decisions to detain suspects during bail hearing (Gazal-Ayal & Sulitzeanu-Kenan, 2010). Moreover, Eberhardt and colleagues found a link between the prototypicality of black, death-eligible convicts' facial features and likelihood of receiving a death versus a long prison term during sentencing (Eberhardt et al., 2006). After controlling for details of the crime and mitigating circumstances, those who had more stereotypically African facial features were more likely to receive a death sentence than those who had less stereotypically African facial features. Given the seriousness with which judges and juries review criminal proceedings in general and death-eligible cases, in particular, it is unlikely that judges and juries were acting on explicit biases in these examples.

Health and Medicine. Similar effects are observed in health and medicine, in which there is evidence that social biases contribute to important health decisions and disparities (Betancourt, Green, Carrillo, & Park, 2005; Brach & Fraserirector, 2000; Krieger, et al., 1993; Williams & Jackson, 2005). For example, Green and colleagues (2007) found that doctors with greater implicit racial bias were less likely to recommend an appropriate treatment for coronary heart disease to Black patients than doctors with less implicit racial bias. In another study, mental health professionals with greater implicit bias against mental illness were more likely than those with less implicit bias to see patient pathology that was not there (Peris, Teachman, & Nosek, 2008). Further, Penner and colleagues (Penner et al., 2010) showed that the quality of doctor-

patient interaction could be predicted by the doctor's implicit racial attitudes. Doctors who self-reported little explicit bias, but had strong implicit bias were rated more negatively by Black patients than were other doctors.

Organizational Behavior. The Civil Rights Act of 1964 prohibits discrimination on the basis of race, gender, and several other group classifications. However, enforcement of this law, for the most part, has assumed that those who discriminate do so intentionally. For example, an employer with a strong commitment to equity and diversity may nonetheless evaluate the same qualifications on a resume differently depending on the race of the applicant. In a naturalistic experiment, Bertrand and Mullainathan (2004) submitted resumes for many job postings across a variety of sectors of the workplace. Resumes with White-sounding names were 50% more likely to receive a callback for an interview than the same resumes with Black-sounding names. The researchers speculated that implicit, rather than explicit, racial biases might be responsible for the discrepancy. Indeed, Rooth (2009) replicated the effect in Sweden; resumes with Swedish-sounding names received more callbacks than ones with Arab-sounding names. He then successfully recruited a portion of the HR professionals who had reviewed the resumes to complete a Swedish-Arab evaluative IAT. Those with stronger implicit pro-Swedish evaluations were more likely to have favored the Swedish applicants (see also Agerstrom & Rooth, 2010 for discrimination evidence toward obese people). Similar discriminatory effects occur in audit studies in others settings such as purchasing a car, renting homes, or tipping cab drivers (Ayers et al., 2004; List, 2004; Ondrich et al., 2003).

Civic behavior. In the United States, voters are assigned to voting locations that might include schools, churches, or other public buildings. This assumes that voting location has no effect on voting behavior. However, evidence suggests that voting in a school may lead voters to vote for more funding for education (Berger, Meredith, & Wheeler, 2008). Further evidence shows how simple, subtle primes can affect attitudes and political opinions. For example, the mere presence of an American flag can increase people's motivations for power and aggression (Ferguson, Carter, & Hassin, 2009; Hassin, Ferguson, Shidlovsky, & Gross, 2007). And, simple reminders of cleanliness increase conservative political attitudes (Helzer & Pizarro, 2011). Finally, subliminally priming Christian concepts is sufficient to increase racial prejudice toward African Americans (Johnson, Rowatt, & LaBouff, 2010). The key point from these examples is that judgment is not an act independent of the social context. A person's "real" opinion is not fixed in time and circumstance, it is influenced by the setting in which that opinion, policy position, and vote is generated.

A common policy strategy: Just be objective. Implicit social cognition evidence challenges the common assumption that people can be objective if they decide to do so. This erroneous assumption is prevalent in a variety of practices. For example, many organizations have non-discrimination policies and report that they follow Employment Equal Opportunity Commission (EEOC) guidelines. By and large, these organizations do little more than instruct managers not to discriminate and fail to systematically evaluate whether the instruction is sufficient to avoid discrimination. Indeed, Bertrand and Mullainathan (2004) found that a company's claim to following EEOC guidelines was not associated with any reduction in discrimination against resumes with black sounding names, compared to those with white sounding names.

Akin to declarations of self-objectivity during voir dire in jury selection, Uhlmann and Cohen (2007) found that asking questions like "Are you objective?" and "Do you use the evidence and facts to make decisions?" (to which virtually all people agree) led to *greater*

reliance on gender stereotypes in a subsequent simulated hiring decision. The researchers argued that the self-declaration of “I am objective” affirmed that one’s thoughts arise from objective sources, so whatever thoughts come to mind are acceptable to use. In a replication and extension of Uhlmann and Cohen’s (2007) study, Lindner, Graser, and Nosek (2011) found that presenting standard EEOC guidelines about avoid discrimination prior to making a hiring led the raters to believe that they had successfully avoided using inappropriate selection criteria, but did not change actual biases in social judgment (see Norton et al., 2004 and Uhlmann & Cohen, 2005 for related examples of failing to recognize bias in one’s own behavior). Despite their prevalence, admonitions against discrimination may not be sufficient on their own to prevent decision-makers from acting on their biases because they may not recognize that they have them or that they are having an impact on behavior.

Alternate policy strategies that improve behavioral realism. If intention, self-awareness, and self-control are not sufficient to eliminate unwanted biases from influencing behavior, then how do we ensure that the individual or organizational goals and values are responsible for the judgment or action? We would be delighted to provide a straightforward answer, but the scientific literature is not yet prepared for that. There is, however, sufficient evidence to provide a framework for behaviorally realistic evaluation of policy alternatives.

First, if particular unwanted biases are identified, interventions can be designed to change the biases themselves (see Blair, 2002; Gawronski & Sritharan, 2010; Nosek et al., in press for reviews). As discussed earlier, interventions such as exposure to counter-stereotypes (Dasgupta & Greenwald, 2001) or to explicit beliefs and values of important individuals in a social context (Sinclair, Huntsinger, Skorinko, & Hardin, 2005) can shift implicit evaluations. The main challenges for this strategy are (a) whether the shift will be durable, especially if the social context is temporary, and (b) whether the intervention can shift all of the unwanted biases. If, for example, the goal is evaluate a target based on relevant job skills and not social features, then the intervention would need to affect social biases relating to age, gender, race, sexual orientation, weight, attractiveness, height, and whatever other social categories are likely vary across the targets. No known intervention has such a pervasive impact.

As such, a second strategy is to give decision-makers strategies to prevent the unwanted biases from being activated or influence judgment (Wilson & Brekke, 1994). For example, being motivated to respond without prejudice can be effective at reducing discriminatory behavior (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; see Bartlett, 2009 for a review). Likewise, instead of adopting an “I am objective” mindset, decision-makers might recognize the potential for bias in decision-making and adopt an internal adversarial approach to considering, questioning and ultimately adopting a position. If the decision-maker actively generates reasons against an initially preferred decision (or reasons in favor of an alternative), then he or she may discover assumptions that made the initial decision appear more attractive than justified. However, the influence of biases in judgment and decision-making are notoriously difficult to prevent, even among people that have relevant knowledge and available strategies (Gilovich, Griffin, & Kahneman, 2002), making it unlikely that these will provide a comprehensive solution on their own.

Finally, a third strategy is to design the behavioral or decision-making context so that they unwanted biases can be identified and corrected, or not influential at all. For example, Goldin and Rouse (2000) found evidence that the slow and then rapid increase of hiring women in major orchestras was caused by the introduction of blind auditions during the review process. Adding a screen removed gender information from a critical stage of evaluation. Gender simply

could not influence judgment because the evaluators did not know the target's gender. This approach has many possible applications, even for circumstances in which the evaluator cannot be completely blind to the target's social characteristics. For example, the legal system employs prescriptive standards in sentencing guidelines as a means to reduce subjective biases. Evidence suggests that departure from these standards is an important source of racial disparities in sentencing (Johnson, 2006). At the same time, judges tend to dislike strict standards because they disallow consideration of unanticipated mitigating circumstances. To provide a balanced approach to these competing concerns the judge could be required to write a structured summary of the mitigating circumstances without any identifying cues, social group characteristics, or other irrelevant information about the defendant and case. Other judges – who are thus blind to those features that are irrelevant for sentencing – could provide feedback on the mitigating circumstances and the proposed sentence. To the extent that the structured summary is free from social biases, this expert feedback would provide an external check for the judge to inform the ultimate sentence.

One could even imagine adapting such a process for performance evaluation of employees. The evaluating supervisor cannot easily be blinded to employee identities, but committees of other supervisors could review performance reports or other performance indicators with identifying information redacted. The comparison of identity-conscious and identity-blind evaluations provides a means to identify discrepancies and investigate whether a social bias might be contributing. In summary, if multiple evaluators collectively have access to all information, but individually are blinded to particular pieces of information, then it is possible to identify whether a given piece of irrelevant information has an unintended impact. Such practices may offer opportunities to correct biased processes even when individual evaluators fail to recognize influences on their own decision-making.

Together, reducing bias, providing decision-making strategies, and restructuring the decision-making process provide substantial opportunity for ensuring that the individual or organizational goals and values are responsible for judgment or action. However, the evidence for implicit social cognition influencing behavior is well-ahead of the evidence for strategies to mitigate that influence when it is unwanted. Even so, development of effective interventions will accelerate rapidly as testing of them occurs in real-world contexts. The application attempts (both failures and successes) will inform the basic research enterprise of determining how and when implicit social cognition influences behavior, and how it can be prevented when it is misaligned with values or goals.

Using implicit measures in social policy

The previous section discussed how existing theory and evidence from implicit social cognition may inform social policy. Another possibility is to incorporate implicit measurement methods themselves into social policies; there are several ways to do this, some easier to justify than others.

A company concerned about the effects of implicit bias in performance evaluations and hiring might develop training procedures in which managers complete implicit measures as an educational exercise about unwanted biases and how to avoid them. Another organization may decide to use implicit measures as a selection device for managers; those who show strong implicit biases against social groups that are important to the organization might be less likely to be hired. A prosecuting (or defense) attorney may administer implicit measures to a defendant to assess whether implicit biases are consistent with the purported criminal act. These illustrate

three potential applications of implicit measures in policy: education, selection, and evaluation. We consider each in turn.

Implicit measures as educational devices. Are implicit measures useful as educational or (un)consciousness raising devices? Project Implicit operates a public website (<https://implicit.harvard.edu/>) at which visitors can learn about implicit biases with descriptive educational materials and self-administered implicit measures that provide feedback about performance. Since its introduction in the fall of 1998, more than 13 million tests have been completed. The site retains persistent popularity with the general public, as part of psychology and other social science courses, and among organizations that send employees during education and training efforts. Further, participant evaluations after completing the measures at the website are highly favorable, on average (Frazier, Nosek, Greenwald, & Banaji, 2010; see also Morris & Ashburn-Nardo, 2010; Morris, Ashburn-Nardo, & Padgett, in press). These minimally suggest that people perceive the experience of completing implicit measures to be educational, perhaps provocative, and, at least, engaging.

However, does experience with implicit measurement have any impact on subsequent beliefs or behavior? Some research suggests that it does. Monteith and colleagues found that experience with an IAT measuring racial bias increased subsequent liking of black names to the extent that the person felt negatively about their implicit bias (Monteith, Ashburn-Nardo, Voils, & Czopp, 2002). Further, Menatti, Smyth, Teachman and Nosek (2011) found that completing an implicit measure of attitudes toward people with mental illness reduced subsequently self-reported stigmatizing mental illness attitudes. However, these studies did not investigate whether such effects were durable beyond the immediate impact of the learning episode. Joy-Gaba and Nosek (2011) evaluated the immediate and long-term impact on beliefs about the existence and operation of implicit bias following a one-hour interactive lecture. The lecture was part of college orientation for the more than 3,000 students in the University of Virginia class of 2014; it included experiential illustrations of automaticity as well as group demonstrations of the IAT. Participants' beliefs about bias and motivation to address bias changed immediately following the presentation, and that change was durable at a follow-up assessment 2-4 months later. However, there was no evidence at the follow-up session that the education mitigated biases in social judgment (Joy-Gaba & Nosek, 2011).

Education using implicit measures might counter the biasing effects of the self-objectivity mindset discussed in the previous section (Uhlmann & Cohen, 2007; Lindner et al., 2011). Direct experience with one's own implicit biases may initiate "cues for control", lead people to monitor their decisions more carefully, and initiate corrective efforts when unwanted thoughts are noticed (Monteith, et al., 2002; Monteith & Mark, 2005; Monteith, Mark, & Ashburn-Nardo, 2010; see also Legault, Green-Demers, & Eadie, 2009). While relatively sparse, the present evidence suggests that using implicit measures as an educational device can increase awareness and knowledge about implicit biases. However, it is not yet clear whether education alone can eliminate the effects of social biases in judgment and behavior.

Implicit measures as selection devices. Educating potential jurists or HR managers with implicit measures is quite different than using performance on implicit measures as an eligibility criterion for selecting jurists or managers. The temptation to pursue implicit measures for selection purposes is easy to understand. As demonstrated by Rooth (2009), HR staff members' implicit racial biases predicted discrimination in receiving job interviews. These results appear to offer a straightforward policy implication: measure HR staff applicants and hire those with the weakest unwanted implicit biases. However, our view is that such an action

would overapply a stereotype: that people with stronger implicit preferences for one social group over others are more likely to behave in biased ways that favor that social group. The reviewed evidence shows that this stereotype is true in the aggregate, but, in our view, applying it to individual cases disregards too much uncertainty in measurement and predictive validity. First, individual scores are sensitive to the context of measurement and show both trait-like and state-like components, so selection based on a single assessment is constrained to the circumstances of assessment. Second, while less fakeable than self-report measures, implicit measures are not immune to deliberate alteration (for a review of studies see Nosek et al., 2007); some progress in identifying faking has not completely removed the possibility of deliberate alteration (Cvencek et al., 2010). Third, the circumstances of predictive validity are not yet well-enough understood to anticipate when a particular implicit bias will influence behavior or how much. Finally, and most importantly, implicit measures assess thoughts, not behaviors - at least not legally actionable behaviors. Social policies are for regulating behaviors, not thoughts. Thinking particular thoughts might increase the likelihood of particular behaviors, but it does not guarantee that they will occur. People who use effective control strategies may never show behavioral evidence of their implicit biases. Likewise, someone with detectable but weak implicit biases may still evidence behavioral effects if he has ineffective control strategies.

Taken together, there is substantial risk for both falsely identifying people as eventual discriminators and failing to identify people who will discriminate. It is possible that, with time, the uncertainties of implicit measurement, predictive validity, and control will decrease sufficiently to make such applications more palatable. However, our own view, which we recognize as our normative position, is that the present sensitivity of implicit measures does not justify this application.

Implicit measures as evaluation devices. Another potential application of implicit measures is the counterpart to selection. Whereas selection policies would use implicit measures to make decisions as an anticipation of future events (would this person be biased or not in the future?), evaluation policies would use implicit measures to determine something about the past. For example, in the legal context, is it likely that a defendant's crime can be classified as a "hate crime" because of implicit negativity toward one of the victim's group memberships?

Many of the same considerations for selection apply for evaluation in the individual case. Making decisions about individuals based on their scores on implicit measures is fraught with uncertainty about the measurement itself and its applicability to the behavior in question. The presence or absence of implicit bias may be associated with particular behavioral outcomes, but there is uncertainty for any individual case; the absence of implicit bias does not guarantee that a behavior was *not* motivated by social biases, and the presence of implicit bias does not guarantee that a behavior *was* motivated by social biases.

Are there any applications to individuals for which the use of implicit measures is justified? Despite our positive disposition about the validity of implicit measures and the use of implicit social cognition theory and evidence in policy, our preceding discussion conveys clear skepticism about the application of implicit measures to *individual* cases in policy.¹⁰ We do not, however, rule out that such applications could be justified eventually. Moreover, we believe that there are some cases that are easier to justify, even without additional evidence. In particular, applying implicit measures to individual cases might be justified when the following criteria are true. First, implicit measures are used as *contributing* evidence, not definitive evidence. In such

¹⁰ Applications to classes of people carry somewhat different considerations, but we do not develop that here.

cases, implicit measures would be considered in combination with many other forms of evidence, such as – in the case of clinical practice – past behavior, structured clinical interview, clinician judgment, and patient self-assessment. Second, the behavior itself is important to predict but challenging to predict with other measures and the implicit measure shows predictive validity in that circumstance. And, third, the relative tradeoffs between false positives and false negatives are acutely in favor of pursuing every means of predictive power.

Few behaviors fit these criteria. The most promising application of this type, in our view, is the use of implicit measures for prediction of acute psychiatric risk, particularly likelihood of committing suicide. Predicting suicidal behavior is difficult (Busch, Fawcett, & Jacobs, 2003; Nock & Banaji, 2007). Nock and colleagues have accumulated evidence that implicit measures associating, for example, the self with life versus death, predict intent to harm the self and, more importantly, future suicide attempts (Nock & Banaji, 2007a, 2007b; Nock et al., 2010; see also Glashouwer & de Jong, 2010). In particular, “the implicit association of death/suicide with self was associated with an approximately 6-fold increase in the odds of making a suicide attempt in the next 6 months, exceeding the predictive validity of known risk factors (e.g., depression, suicide-attempt history) and both patients’ and clinicians’ predictions” (Nock et al., 2010, p. 511). Presently, this holds the most promise for implicit measures assisting decision-making in individual cases, because (a) many sources of evidence are pooled in consideration of suicide risk and acute clinical care; (b) existing measures for identifying suicide risk are relatively ineffective; (c) evidence suggests that implicit measures add substantial predictive accuracy for suicide attempts; and (d) the tradeoff between taking extra precautions for a person incorrectly identified as being at risk for suicide versus reducing the risk that someone fails to be identified as at risk and then commits suicide is strongly tilted toward the latter.

As with all cumulative science, the relevant evidence that informs the merits of using implicit measures may change as the evidence accumulates. Ultimately, any application of psychological measures in policy is a value judgment based on an assessment of the tradeoffs and goals of the policy. Scientific evidence can contribute to informing about the likely (or unlikely) success of the policies, but it cannot make irrelevant the use of values to decide what *should* be done.

References

- Agerström, J., & Rooth, D.-O. (2011). The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology*. Advance online publication. doi:10.1037/a0021594
- Arkes, H. & Tetlock, P. E. (2004) Attributions of implicit prejudice, or “Would Jesse Jackson ‘fail’ the Implicit Association Test?” *Psychological Inquiry*, 15, 257- 278.
- Ayers, I., Vars, F. E., & Zakariya, N. (2004). To insure prejudice: Racial disparities in taxicab tipping. *Yale Law Journal*, 114, 1613-1674.
- Bagenstos, S.R. (2007). Implicit bias, ‘science’, and antidiscrimination law. *Harvard Law and Policy Review*, 1, 477-493.
- Banaji, M. R. (2001). Ordinary prejudice. *Psychological Science Agenda, American Psychological Association*, 14 (Jan-Feb), 8-11.
- Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes and Tetlock. *Psychological Inquiry*, 15, 279-310.
- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of implicit racial attitudes. *Journal of Personality and Social Psychology*, 87, 5–22.
- Bargh, J. A. (1994) The Four Horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer, Jr., and T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40) Hillsdale, NJ: Erlbaum.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462-479.
- Bartlett, K. T. (2009). Making good on good intentions: The critical role of motivation in reducing implicit workplace discrimination. *Virginia Law Review*, 95, 1893-1972.
- Bazelon, D. L. (1982). Veils, values, and social responsibility. *American Psychologist*, 37, 115-121.
- Bem, D. J. (1972) Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, (Vol. 6, pp. 1-62) New York: Academic Press.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Brendan more employable than Lakisha and Jamal? *The American Economic Review*.
- Betancourt, J. R., Green, A. R., Carrillo, J. E., & Park, E. R. (2005). Cultural competencies and health care disparities: Key perspectives and trends. *Health Affairs*, 24, 499-505.
- Bessenoff, G. R. & Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: Automatic evaluation and stereotyping. *Social Cognition*, 18, 329-353.
- Blair, I.V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242 -261.
- Blanton, H., Jaccard, J., Gonzales, P.M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, 42, 192-212.
- Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94, 567–582.
- Blasi, G., & Jost, J. (2006). System justification theory and research: Implications for law, legal advocacy, and social justice. *California Law Review*, 94, 1119–1168.

- Bosson, J.K., Swann, W.B., & Pennebaker, J.W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631-643.
- Brach, C., & Fraserirector, I. (2000). Can cultural competency reduce racial and ethnic health disparities? A review and conceptual model. *Medical Care Research Review*, *57*, suppl 181-217.
- Busch, K. A., Fawcett, J., & Jacobs, D. G. (2003). Clinical correlates of inpatient suicide. *Journal of Clinical Psychiatry*, *64*, 14-19.
- Cai, H., Sriram, N., Greenwald, A. G., & McFarland, S. G. (2004). The Implicit Association Test's D measure can minimize a cognitive skill confound: Comment on McFarland and Crouch (2002). *Social Cognition*, *22*, 673-684.
- Chaiken, S. and Y. Trope (1999). *Dual-process theories in social psychology*. New York, Guilford Press.
- Chugh, D. (2004). Societal and managerial implications of implicit social cognition: Why milliseconds matter. *Social Justice Research*, *17*, 203-222.
- Conrey, F.R., Sherman, J.W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469-487.
- Cvencek, D., Greenwald, A. G., Brown, A., Snowden, R., Gray, N. (2010). Faking of the Implicit Association Test is statistically detectable and partly correctable. *Basic and Applied Social Psychology*, *32*, 302-314.
- Cunningham, W.A., Preacher, K.J., & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163-70.
- Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, *30*, 1332-1346.
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary-school children. *Child Development*, *82*, 766-789.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800-814.
- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*, 347-368.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18.
- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*, 835-848.
- Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*, 286-303.
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383-386.
- Fazio, R.H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M.P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229-238.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (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.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297-327.
- Ferguson, M. J., Carter, T. J., & Hassin, R. R. (2009). On the automaticity of nationalist ideology: The case of the USA. In J. T. Jost, A. C. Kay, & H. Thorisdottir (Eds.), *The Social and Psychological Bases of Ideology and System Justification* (pp. 53-84). Oxford, UK: Oxford University Press.
- Frazier, R.S., Nosek, B.A., Banaji, M.R., Greenwald, A.G. (2010). "Participants' responses to completing the race Implicit Association Test." Poster presented at the Society for Personality and Social Psychologists annual conference, Las Vegas, NV.
- Freud, S. (1900/1972). *Interpretation of dreams*. New York: Basic Books.
- Fulero, S. M. (1986). The role of behavioral research in the free press/fair trial controversy: Another view. *Law and Human Behavior*, *11*, 259-264.
- Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science*, *321*, 1100-1002.
- Gardner, W., Scherer, D., & Tester, M. (1989). Asserting scientific authority: Cognitive development and adolescent legal rights. *American Psychologist*, *44*, 895-902.
- Gawronski, B., & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692-731.
- Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin*, *34*, 648-665.
- Gawronski, B., & Payne, B.K. (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford Press.
- Gawronski, B., & Sritharan, R. (2010). Formation, change, and contextualization of mental associations: Determinants and principles of variations in implicit measures. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 216-240). New York: Guilford Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press.
- Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of 'blind' auditions on female musicians," *American Economic Review*, *90*, 715-741.
- Goodman, G. S., Levine, M., & Melton, G. (1992). The best evidence produces the best law. *Law and Human Behavior*, *16*, 244-251.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, *22*, 1231-1238.
- Greenwald, A. G. (1992). New Look 3: Reclaiming unconscious cognition. *American Psychologist*, *47*, 766-779.

- Greenwald, A.G. (2008). Landy is correct: Stereotyping can be moderated by individuating the out-group and by being accountable. *Industrial and Organizational Psychology, 1*, 430-435.
- Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.
- Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022-1038.
- Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.
- Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review, 94*, 945-967.
- Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the Implicit Association Test: Comment on the article by Blanton and Jaccard. *American Psychologist, 61*, 56-61.
- Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean? In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the New Implicit Measures* (pp. 65-82). Hillsdale, NJ: Erlbaum.
- Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009) Race attitude measures predicted vote in the 2008 U. S. Presidential Election. *Analyses of Social Issues and Public Policy, 9*, 241-253
- Griffin, T., & Wooldredge, J. (2006). Sex-based disparities in felony dispositions before versus after sentencing reform in Ohio. *Criminology, 44*, 893-923.
- Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally-created extrapersonal associations on the Implicit Association Test. *Journal of Experimental Social Psychology, 42*, 259-272.
- Hassin, R. R., Ferguson, M. J., Shidlovsky, D., & Gross, T. (2007). Waved by invisible flags: The effects of subliminal exposure to flags on political thought and behavior. *Proceedings of the National Academy of Sciences, 104*, 19757-19761.
- Helmholtz, H. (1925). *Treatise on physiological optics* (3rd ed., Vol. 3; J.P.C. Southall, Trans.). Menasha, WI: Banta. (Original 3rd ed. published 1910).
- Helzer, E. & Pizarro, D.A. (2011) Dirty Liberals!: Reminders of cleanliness promote conservative political and moral attitudes. *Psychological Science, 22*, 517-522.
- Iyengar, S., Huberman, G, & Jiang, W. (2004). How much choice is too much? Contributions to 401(k) retirement plans. In Mitchell, O. S. & Utkus, S. (Eds.) *Pension Design and Structure: New Lessons from Behavioral Finance*, 83-95. Oxford: Oxford University Press.
- Iyengar, S.S., & Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79*, 995-1006.
- Jolls, C., & Sunstein, C. R. (2006). The law of implicit bias. *California Law Review, 94*, 969-996.
- Johnson, B. D. (2003). Racial and ethnic disparities in sentencing: Departures across modes of conviction. *Criminology, 2003*, 449-490.

- Johnson, M. K., Rowatt, W. C., & LaBouff, J. (2010). Priming Christian religious concepts increases racial prejudice. *Social Psychological and Personality Science, 1*, 119-126.
- Jost, J.T. (2006). The end of the end of ideology. *American Psychologist, 61*, 651-670.
- Jost, J.T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior, 29*, 39-69.
- Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology, 41*, 137-146.
- Kang, J. (2006). Trojan horses of race. *Harvard Law Review, 118*, 1489-1593.
- Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of "Affirmative Action." *California Law Review, 94*, 1063-1118.
- Kang, J. & Lane, K. (2009). Seeing through colorblindness: Implicit bias and the law. *UCLA Law Review, 58*, 465-521.
- Karpinski, A., & Hilton, J.L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774-788.
- Karpinski, A., Steinman, R. B., & Hilton, J. L. (2005). Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. *Personality and Social Psychology Bulletin, 31*, 949-962.
- Kassin, S. & Sommers, S. (1997). Inadmissible testimony, instructions to disregard, and the jury: Substantive versus procedural considerations. *Personality and Social Psychology Bulletin, 23*, 1046-1054.
- Kassin, S. & Studebaker, C. A. (1998). Instructions to disregard and the jury: Curative and paradoxical effects. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 413-452). Hillsdale, NJ: Lawrence Erlbaum.
- Kitchell, J. F., Martell, S. J. D., Walters, C. J., Jensen, O. P., Kaplan, I. C., Watters, J., Essington, T. E., & Boggs, C. H. (2006). Billfishes in an ecosystem context. *Bulletin of Marine Science, 79*, 669-682.
- Krieger, L. H., & Fiske, S. T. (2006). Behavioral realism in employment discrimination law: Implicit bias and disparate treatment. *California Law Review, 94*, 997-1062.
- Krieger, N., Rowley, D. L., Herman, A. A., Avery, B., et al. (1993). Racism, sexism, and social class: Implications for studies of health, disease, and well-being. *American Journal of Preventive Medicine, 9* (6, Suppl), 82-122.
- Kunst-Wilson, W. & Zajonc, R. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207*, 557-558.
- Landy, F.J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology, 1*, 379-392.
- Leavitt, K., Fong, C.T., & Greenwald, A.G. (2011). Asking about well-being gets you half an answer: Intra-individual processes of implicit and explicit job attitudes. *Journal of Organizational Behavior, 32*, 672-687.
- Legault, L., Green-Demers, I., & Eadie, A.L. (2009). When internalization leads to automatization: The role of self-determination in automatic stereotype suppression and implicit prejudice regulation. *Motivation and Emotion, 33*, 10-24.
- Lee, S., Rogge, R. D., & Reis, H. T. (2010). Assessing the seeds of relationship decay: Using implicit evaluations to detect the early stages of disillusionment. *Psychological Science, 6*, 857-864.

- Lindner, N.M., & Nosek, B.A. (2009). Alienable speech: Ideological variations in the application of free-speech principles. *Political Psychology, 30*, 67-92.
- Lindner, N. M., Nosek, B. A., & Graser, A. (2011). Age-based hiring discrimination as a function of equity norms and self-perceived objectivity. Unpublished manuscript.
- List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *Quarterly Journal of Economics, 119*, 49-89.
- Martin, L. L., & Tesser, A. (1992). *The Construction of Social Judgments*. New York: Psychology Press.
- McConnell, A.R., & Leibold, J.M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435-442.
- McConnell, A.R. & Leibold, J.M. (2009.) Weak criticisms and selective evidence: Reply to Blanton et al. (2009). *Journal of Applied Psychology, 94*, 583-589.
- McFarland, S.G. & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 6*, 483-510.
- Mierke, J., & Klauer, K.C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*(6), 1180-1192.
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80*, 252-283. doi: [10.1037/h0035002](https://doi.org/10.1037/h0035002)
- Mischel, W. & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*, 246-268. doi: [10.1037/0033-295X.102.2.246](https://doi.org/10.1037/0033-295X.102.2.246)
- Mitchell, J.P., Nosek, B.A., & Banaji, M.R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*, 455-469.
- Mitchell, G. & Tetlock, P.E. (2006). Antidiscrimination law and the perils of mindreading. *Ohio State Law Journal, 67*, 1023-1121.
- Molesworth, B. R. C., & Chang, B. (2009). Predicting pilots' risk-taking behavior through an Implicit Association Test. *Human Factors, 51*, 845-857.
- Monteith, M.J., Ashburn-Nardo, L., Voils, C.I., & Czopp, A.M. (2002). Putting the brakes of prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology, 83*, 1029-1050.
- Monteith, M.J., & Mark, A.Y. (2005). Changing one's prejudiced ways: Awareness, affect, and self-regulation. *European Review of Social Psychology, 16*, 113-154.
- Monteith, M.J., Mark, A.Y., & Ashburn-Nardo, L. (2010). The self-regulation of prejudice: Toward understanding its lived character. *Group Processes and Intergroup Relations, 13*, 183-200.
- Monteith, M.J., Voils, C.I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition, 19*, 395-417.
- Mook, D.G. (1983). In defense of external invalidity. *American Psychologist, 38*, 379-87.
- Moors, A. and De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297-326.
- Morris, K. A., & Ashburn-Nardo, L. (2010). The Implicit Association Test as a class assignment: Student affective and attitudinal reactions. *Teaching of Psychology, 37*, 63-68.
- Morris, K. A., Ashburn-Nardo, L., & Padgett, R. J. (in press). Think fast: Using web-based reaction time technology to promote teaching about racial bias and diversity. Chapter to

- appear in D. S. Dunn, J. C. Wilson, J. Freeman, & J. R. Stowell (Eds.), *Getting connected: Best practices for technology-enhanced teaching and learning in psychology*. Oxford University Press.
- Mueller, C. B., & Kirkpatrick, L. C. (1995). *Modern evidence: Doctrine and practice*. Boston: Little, Brown.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64*, 723-739.
- Mustard, D. B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the U.S. federal courts. *Journal of Law and Economics, 44*, 285-314.
- Nisbett, R.E., & Wilson, T.D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*, 250-256.
- Nock, M. K., & Banaji, M. R. (2007a). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Consulting and Clinical Psychology, 75*, 707-715.
- Nock, M. K., & Banaji, M. R. (2007b). Assessment of self-injurious thoughts using a behavioral test. *American Journal of Psychiatry, 164*, 820-823.
- Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science, 21*, 511-517.
- Nosek, B.A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565-584.
- Nosek, B.A. (2007). Understanding the individual implicitly and explicitly. *International Journal of Psychology, 42*, 184-188.
- Nosek, B.A. & Banaji, M.R. (2001). The go/no-go association task. *Social Cognition, 19*, 625-666.
- Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Math = male, me = female, therefore math \neq me. *Journal of Personality and Social Psychology, 83*, 44-59.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36-88.
- Nosek, B.A., & Smyth, F.L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology, 54*, 14-29.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265-292). New York: Psychology Press.
- Nosek, B.A., & Hansen, J.J. (2008). Personalizing the Implicit Association Test increases explicit evaluation of target concepts. *European Journal of Psychological Assessment, 24*, 226-236.
- Nosek, B.A., & Greenwald, A.G. (2009). (Part of) the case for a pragmatic approach to validity: Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin, 135*, 373-376.

- Nosek, B.A., Smyth, F.L., Sriram, N., Lindner, N.M., Devos, T., Ayala, A., Bar-Anan, Y., et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*, 10593–10597.
- Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, *15*, 152–159.
- Olson, M.A., & Fazio, R.H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.
- Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, *14*, 636–639.
- Ondrich, J., Ross, S., & Yinger, J. (2003). Now you see it, now you don't: Why do real estate agents withhold available houses from black customers? *The Review of Economics and Statistics*, *85*, 854–873.
- Payne, B.K., Cheng, C.M., Govorun, O., & Stewart, B.D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293.
- Payne, B. K., McClernon, F. J., & Dobbins, I.G. (2007). Automatic affective responses to smoking cues. *Experimental and Clinical Psychopharmacology*, *15*, 400–409.
- Payne, B.K., Krosnick, J.A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology*, *46*, 367–374.
- Penner, L. A., Dovidio, J. F., West, T. V., Gaertner, S. L., Albrecht, T. L., Dailey, R. K., & Markova, T. (2010). Aversive racism and medical interactions with Black patients: A field study. *Journal of Experimental Social Psychology*, *46*, 436–440.
- Peris, T. S., Teachman, B. A., & Nosek, B. A. (2008). Implicit and explicit stigma of mental illness: Links to clinical care. *Journal of Nervous and Mental Disease*, *196*, 752–760.
- Petróczi, A., Aidman, E. V., Hussain, I., Deshmukh, N., Nepusz, T., Uvacek, M., Tóth, M., Barker, J., & Naughton, D. P. (2010) Virtue or pretense? Looking behind self-declared innocence in doping. *PLoS ONE*, *5*, e10457. doi:10.1371/journal.pone.0010457
- Rachlinski, J. J., Johnson, S.L., Wistrich, A.J., & Guthrie, C. (2009). Does unconscious racial bias affect trial judges? *Notre Dame Law Review*, *84*, 09–11.
- Ranganath, K.A., Smith, C.T., & Nosek, B.A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44*, 386–396.
- Ranganath, K. A. & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately, explicit attitude generalization takes time. *Psychological Science*, *19*, 249–254.
- Roccatò, M., & Zogmaister, C. (2010). Predicting the vote through implicit and explicit attitudes: A field research. *Political Psychology*, *31*, 249–274.
- Rooth, D-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, *17*, 523–534.
- Rothermund, K. & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*, 139–165.
- Rudman, L. A. (2008). The validity of the Implicit Association Test is a scientific certainty. *Industrial and Organizational Psychology*, *1*, 426–429.

- Rydell, R.J. & McConnell, A.R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*, 995-1008.
- Schauer, F. (2009). Can bad science be good evidence: Lie detection, neuroscience, and the mistaken conflation of legal and scientific norms. *Cornell Law Review*, *95*, 1191-1219.
- Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66.
- Schwartz, M. B., Vartanian, L. R., Nosek, B. A., & Brownell, K. D. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity*, *14*, 440-447.
- Schwarz, N. (1999) Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93-105.
- Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz (Eds.), *Blackwell Handbook of Social Psychology: Intraindividual Processes* (pp. 436-457). Oxford, UK: Blackwell.
- Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 156-175). New York: Guilford Press.
- Sherman, S.J., Rose, J.S., Koch, K., Presson, C.C., Chassin, L. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology*, *22*, 13-39.
- Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127-190.
- Sinclair, S., Huntsinger, J., Skorinko, J., & Hardin, C. D. (2005). Social tuning of the self: Consequences for the self-evaluations of stereotype targets. *Journal of Personality and Social Psychology*, *89*, 160-175. doi: 10.1037/0022-3514.89.2.160
- Sinclair, S., Lowery, B.S., Hardin, C.D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative orientation. *Journal of Personality and Social Psychology*, *89*, 583-592.
- Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108-131.
- Sriram, N., Greenwald, A. G., & Nosek, B. A. (2010). Correlational biases in mean response latency differences. *Statistical Methodology*, *7*, 277-291.
- Srivastava, S., & Banaji, M. R. (2011). Behind the front: Collaborative networks and implicit cognition in organizations. *American Sociological Review*, *76*, 207-233.
- Teachman, B. A., Cody, M. W., & Clerkin, E. M. (2010). Clinical applications of implicit social cognition theories and methods. In B. Gawronski and K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 489-521). New York: Guilford Press.
- Teige-Mociemba, S., Klauer, K.C., & Sherman, J.W. (2010). A practical guide to Implicit Association Tests and related tasks. In B. Gawronski and B. K. Payne (Eds.), *Handbook of Implicit Social Cognition* (pp. 117-139) New York, NY: Guilford.

- Tetlock, P. & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior*, 29, 3-38.
- Teubel, T., Asendorpf, J. B., Banse, R., & Schnabel, K. (in press). Implicit but not explicit aggressiveness predicts performance outcome in basketball players. *International Journal of Sport Psychology*.
- Tierney, J. (2008, November 7). Where have all the bigots gone? *New York Times*, retrieved from www.nytimes.com.
- Underwager, R. & Wakefield, H. (1992). Poor psychology produces poor law. *Law and Human Behavior*, 16, 233-243.
- Uhlmann, E.L., & Cohen, G.L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, 16, 474-480.
- Uhlmann, E., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, 104, 207-223.
- Uhlmann, E. L., Poehlman, T. A., & Nosek, B. A. (in press). Automatic associations: Personal attitudes or cultural knowledge?. In J. Hanson (Ed.), *Ideology, Psychology, and Law*. Oxford, UK: Oxford University Press.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wentura, D. & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski and B.K. Payne (Eds.), *Handbook of Implicit Social Cognition* (pp. 95-116). New York, NY: Guilford.
- Wiers, R.W., Rinck, M., Kordts, R., Houben, K., Strack., F. (2010). Retraining automatic action-tendencies to approach alcohol in hazardous drinkers. *Addiction*, 105, 279-287.
- Williams, D. R., & Jackson, P. B. (2005). Social Sources of Racial Disparities in Health. *Health Affairs*, 24, 325-334.
- Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.
- Wilson, T. D. and Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116, 117-142.
- Zajonc, R.B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1-27.
- Ziegert, J. C., & Hanges, P. J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, 90, 553-562.
- Zeigart, J.C. & Hanges, P.J. (2009). Strong rebuttal for weak criticisms: Reply to Blanton et al. (2009). *Journal of Applied Psychology*, 94, 590-597.