

Divided Minds, Divided Morals: How Implicit Social Cognition Underpins and Undermines Our Sense of Social Justice

B. Keith Payne and C. Daryl Cameron

University of North Carolina at Chapel Hill

*Virtue is a state of war, and to live in it we have
always to combat with ourselves.*

Jean-Jacques Rousseau

Of all the reasons people study implicit social cognition, a concern with its implications for social justice probably ranks near the top. The message of implicit social cognition is that the thoughts people introspect and report about do not tell the whole story of why they believe the things they believe, and why they do the things they do. Instead, many studies have found that implicit reactions are prone to prejudice and stereotyping, even when explicit responses are fair-minded. Implicit reactions are often hierarchical, dividing the world into “us” and “them,” even when explicit responses are inclusive. Implicit reactions are impulsive, even when explicit responses are calculated. Implicit cognition can be, to paraphrase Hobbes, nasty, brutish, and short-sighted. If implicit cognition is a driving force behind our personal choices and public actions, then it offers a new way to understand many puzzles in modern social life. Why do people perpetuate inequalities when they apparently mean to be fair? Why do people on the other side of culture war debates seem so certainly wrong, yet so wrongly certain?

In this essay we survey empirical research in implicit social cognition with a view toward what it means for questions of social justice and moral concerns. We begin by examining empirical studies of implicit prejudice, which have clear implications for social inequality and social justice. We then consider a broader set of phenomena, in which implicit social cognition shapes the very judgments of what is just and moral in the first place. We end by considering the ways that implicit social cognition can shed light on processes leading to both just and unjust outcomes, and the role of

empirical research in addressing these complex and value-laden questions.

Implicit Prejudice and Social Justice

Research on implicit prejudice has advanced quickly. This fast pace has sparked lively debate over methods and interpretations of the research. We focus here on interpretations and conclusions that can be drawn from this research, as opposed to methodological issues. Methodological questions are addressed at length in other chapters (De Houwer & Moors, this volume; Klauer, Teige-Mocigemba, & Sherman, this volume; Sekaquaptewa, von Hippel, & Vargas, this volume; Wentura & Degner, this volume).

Subtle behavioral effects of implicit bias

Early studies demonstrated that implicit measures were associated with certain specific aspects of behavior in cross-race interactions (Dovidio et al., 2002; Fazio et al., 1995; McConnell & Leibold, 2001). For instance, implicit tests completed by White subjects were shown to predict their nonverbal *micro-behaviors* toward a Black interaction partner. These micro-behaviors included things such as eye contact, seating distance, and speech errors. Explicit measures of race attitudes did not predict these behaviors, but instead predicted the favorability of explicit ratings about the interaction partner.

Consider a scenario based on this research. John is White and James is Black, and the two interact with the best of intentions. It turns out that John is implicitly biased. His bias reveals itself through subtle micro-behaviors such as poor eye contact, more blinking, speech errors, and less nonverbal friendliness. John is not fully aware of these behaviors, but James picks up on them, and

judges John to be less friendly and trustworthy because of them. If John has a stereotype that Blacks are aggressive, it might subtly influence his behavior toward James, who might respond with his own hostile reaction (Chen & Bargh, 1997). Because John is unaware of the ways in which implicit stereotypes influence nonverbal behavior, he will perceive the hostile behavior from James as coming from nowhere, and this will further solidify the behavioral stereotype about Blacks being aggressive (Word, Zanna, & Cooper, 1974). This kind of self-fulfilling prophecy suggests that micro-behaviors toward outgroup members can cumulatively build a hostile intergroup environment. Because of poor introspection over attitudes and behavior, people will feel as if they have acted objectively and free from bias, and any challenge to that perception will only increase antagonism (Kennedy & Pronin, 2008).

Aside from building distrust and hostility, this environment may harm the performance of those who are biased against. Salvatore and Shelton (2007) had both Black and White participants read resumes for fictional job applicants (some of whom were Black), followed by a decision and justification by a human resources officer. There were three kinds of decision: non-discriminatory, ambiguous (officer chose a lower-qualified ingroup member over a higher-qualified outgroup member, but did not specify reasons), and blatant (same choice, but race was given as the reason). On a Stroop task – which involves using controlled cognitive resources to inhibit the effect of word color on naming color words – White participants performed the worst when exposed to blatant prejudice, whereas Black participants performed the worst when exposed to *ambiguous* prejudice. This finding suggests that hostile environments clouded by implicit bias predict poorer cognitive performance among minority members.

This finding is reminiscent of the large literature on stereotype threat. In such studies, Black students often perform worse than White students on a test when it is said to be diagnostic of intelligence and when racial group membership is made salient (Steele, 1997). This research suggests that to the

extent that an environment is made hostile through minority members' awareness of bias around them, that environment might in turn confirm the very stereotypes that feed the biases. Kang and Banaji (2006) have argued that studies like these throw doubt upon a truly objective criterion for merit in the workplace.

In light of this, a number of studies have investigated how people make decisions based upon “objective” criteria like merit, only to mask their subjective biases. Norton, Vandello, and Darley (2004) investigated the degree to which people would justify hiring a man over a woman for a stereotypically male job (construction foreman). Among a pack of resumes, two stood out as the best, but on different dimensions: one candidate had more education than experience, the other more experience than education. These two resumes were randomly assigned to either male or female candidates. Participants selected male candidates the majority of the time. When asked to justify their decisions, participants in all but the female-highly-educated condition appealed to the education qualification. In this last condition, participants appealed to the experience qualification instead. Subsequent studies suggested that this public casuistry might extend into private consciousness as well, because casuistry-based decisions were shown to impair memory of the very criteria used to make the decision.

Together, these studies make some striking claims about the ways that subtle automatic biases can pose a challenge for social justice. First, implicit bias predicts micro-level behaviors that leak out during interracial interactions. These behaviors tend to be perceived by observers, potentially feeding a cycle of behavioral confirmation and intergroup conflict. Second, an environment perceived as biased leads to performance deficits, in terms of reduced cognitive performance and stereotype threat. Finally, when actions and decisions are driven by subtle biases they may be masked using objective criteria, flexibly deployed to rationalize the decision. The foregoing research has been used by some psychologists and legal scholars to push for large-

scale institutional and legal reform (Kang, 2005; Kang & Banaji, 2006; Krieger & Fiske, 2006; Lane et al., 2007). But these conclusions have not gone unchallenged. We next consider some critiques of the implicit bias literature in light of whether it really supports conclusions that challenge social justice. Then we will discuss recent research that has been stimulated, in part by this critique, to examine behaviors with greater external validity and direct societal implications.

Should we call it prejudice?

Some scholars have argued that as researchers have probed for more and more subtle forms of bias, they have set too low a threshold for what counts as prejudice. Arkes and Tetlock (2004; also Mitchell & Tetlock, 2007) raised a number of criticisms of implicit bias research on both methodological and conceptual grounds. Some of these criticisms were directed specifically at the Implicit Association Test (IAT; Greenwald et al., 1998), and may not apply to the broader literature on implicit bias, which has used multiple methods. We focus here on those criticisms that directly concern whether demonstrations of implicit bias should be interpreted to reveal prejudice, and whether those demonstrations are relevant for social justice.

One criticism is that implicit biases in which minorities are more easily associated with negative than positive concepts could result not only from disliking the minority group, but also from other affective reactions that would not count as prejudice. For example, Arkes and Tetlock (2004) point out that an implicit test might detect bias “against” Blacks if the respondent simply feels nervous about interacting with an outgroup member or feels bad about historical injustice toward Black Americans. Indeed, Uhlmann and colleagues (2006) found that both dislike for a group and feelings of sympathy for the group’s experience of oppression resulted in significant implicit bias on an IAT.

If apparent implicit bias can result from both dislike and sympathy, then there is indeed some ambiguity that needs to be resolved. Yet at first blush, this critique only makes sense given a theory

of emotions that allows for distinct emotions at the automatic level. Some theories suggest that automatic affect is undifferentiated, being qualitatively parsed into discrete emotions only at later stages of processing (Baumeister et al., 2007; Russell, 2003). Should such theories be accurate, Arkes and Tetlock’s emotions-based criticism loses a great deal of force. From this perspective, automatic negativity is simply negativity, with no differences that can be identified as “anxiety” versus “hostility.” Presumably the only basis for judging whether such negativity counts as prejudice would be whether it drives discriminatory treatment, which has already been demonstrated in the studies described above.

Other recent research, however, suggests that discrete automatic emotions can indeed be distinguished. Arbuckle and Payne (2008), for instance, used the Affect Misattribution Procedure (Payne, Cheng, Govorun, & Stewart, 2005) to measure discrete emotional reactions. Participants saw White American or Middle Eastern faces as primes, followed on each trial by a Chinese pictograph. In one phase of the task, participants guessed whether each pictograph meant “fear.” In another phase, they guessed whether each pictograph meant “anger.” The Middle Eastern primes elicited greater fear and anger responses than White American primes. Participants also expressed their opinions about how to deal with American-Iranian tensions over the possibility of Iran developing nuclear weapons. Implicit anger predicted support for offensive military action against Iran, whereas implicit fear predicted non-military defensive action. The critique that multiple emotions may underpin implicit biases is not necessarily a weakness, if researchers realize that multiple emotions may have distinct effects and make an effort to elucidate these effects.

In addition to the point about emotions, Arkes and Tetlock (2004) suggest that implicit race biases present no serious ethical challenge. They argue that these biases are not personally endorsed attitudes, but rather culturally-borne stereotypes that would be activated for any rational person familiar with cultural norms and statistical base rates. How could

researchers then attribute the morally loaded term “prejudice” to a reaction that was not personally endorsed?

First, some authors have suggested that whether not an attitude is personally endorsed is irrelevant if it leads to unjust behavior (Nosek & Hansen, 2008). Moreover, the conceptual boundary of what reflects one’s “true self” is contestable: some suggest that what counts as a person’s “true self” involves only those attitudes that are reflectively endorsed (Frankfurt, 1969), whereas others suggest that the most revealing attitudes are the ones that occur spontaneously and unintentionally (Smith, 2005; see Gawronski et al., 2008 for further discussion of this point).

Second, even if cultural stereotypes are rational in the sense of being statistically defensible (e.g. using base-rates in making social judgments), they might be immoral (Banaji, Nosek, & Greenwald, 2004). Moral claims involve not only whether a belief is statistically grounded, but also whether people are harmed by it. The risk of harm for the holder of a stereotyped belief is very different from the risk of harm for the target of that belief. For example, although a police officer who “racially profiles” motorists might justify differential treatment using base rates, the innocent suspect that is pulled over has his own morally relevant considerations, often discussed in terms of rights. The motorist rightfully expects to be treated as an individual based on his conduct rather than group stereotypes.

The moral relevance from the perspective of the target of prejudice is highlighted by empirical research over the last decade which in our view has clear and startling implications for social justice. Most conceptions of justice argue that all people are entitled to certain rights derived from their shared humanity. History attests to the link between dehumanization and moral license: those who are dehumanized are pushed beyond the scope of rights that would preclude the most egregious atrocities (Bandura, 1999). Studies demonstrating links between implicit social cognition and dehumanization provide striking examples of social cognition’s moral relevance.

One aspect that distinguishes humans from non-human animals is the experience of secondary emotions: those social and self-conscious emotions that tend to be distinctive to human experience. Although many animals may feel fear and rage, for example, it is often presumed that only humans feel guilt or embarrassment. Using an adaptation of the IAT, Leyens and colleagues (2000) found that participants responded faster when ingroup names and secondary emotions were paired than when outgroup names and secondary emotions were paired. Leyens and colleagues interpreted their findings as evidence that people associate out-group members with less human qualities than in-group members.

Even more directly, Goff and colleagues (2008) recently investigated the implicit association between Blacks and apes. To test for the implicit presence of this subjugating metaphor, they primed participants with either Black or White faces and asked them to identify perceptually degraded images of animals as fast as possible. Priming Black faces facilitated the recognition of apes, whereas White faces slowed down this recognition. In another study, participants primed with apes on a dot probe task showed increased attention to Black faces as opposed to White faces. Moreover, this effect did not generalize to Asian faces, but was specific to Black faces.

Using an IAT to measure the association between Blacks and apes, Goff et al. (2008) had participants categorize human names as Black or White, and animals as apes or big cats. Participants were faster to respond when Blacks and apes were paired, and this effect was independent of bias on a traditional evaluative race IAT. Participants also tended to deny explicit knowledge of the Black-ape metaphor. Connecting this to real-world moral behavior, participants were more likely to say that police brutality against Black (v. White) suspects was justified after being primed with apes (vs. big cats). Finally, Goff and colleagues performed a content analysis of the media surrounding death-penalty eligible criminal cases mentioned in the *Philadelphia Inquirer* between 1979 and 1999. Black defendants were described using more ape-

like metaphors and imagery, and this trend was associated with higher rates of capital punishment for Blacks.

Many of the points raised by Arkes and Tetlock (2004) are thus being addressed and incorporated into research on implicit race bias. The critique that various emotional reactions besides antipathy can produce biased scores on implicit measures suggests a valuable caveat in interpreting some implicit bias research. The critique has also been valuable in stimulating new research that distinguishes more carefully between the specific emotional underpinnings of bias. This research shows that multiple emotional responses can have meaningful impacts. Moreover, it shows that some associations revealed by implicit tests go well beyond the simple good-bad evaluations targeted by the critique. It is hard to claim that associations between Black Americans and apes are too mild to be considered prejudice.

Associations between subtle biases and discriminatory behavior provide evidence that these biases have implications beyond the laboratory. In the next section we consider more studies that make clear and direct connections to meaningful overt behavior. These studies highlight further the practical importance of implicit social cognition for social justice, by documenting how implicit bias relates to ethically unambiguous outcomes.

Less-than-subtle behavioral effects of implicit bias

Implicit race bias and criminal justice. Studies of implicit bias in the laboratory often ask respondents to make snap judgments, as in the speeded responses on implicit attitudes tests. In some cases, these are far removed from the kinds of behaviors these tests are intended to model or predict. But in some cases, it is the snap judgment itself that is the meaningful behavior. Consider the case of Amadou Diallo, the African immigrant whose wallet was mistaken for a gun by New York police officers. This mistake led to Diallo being shot to death, and subsequent questions revolved around the degree to which any bias on the police officers' part influenced their perceptions and decisions to shoot. Moral outrage in response to this incident derived from a salient principle of justice:

that an individual should be judged on the basis of his conduct, not his social category.

In the wake of this incident, a number of studies examined the psychological underpinnings of race biases that could lead to such a mistake. In one study, subjects were asked to identify an object as either a gun or a harmless hand tool (Payne, 2001). Just prior to seeing each object, participants saw either a Black or White face. When making snap judgments, subjects were more likely to make errors in saying that a harmless object was a gun after being primed with a Black face. This weapon bias did not depend on intentional bias; in fact it occurred in spite of intentions to the contrary. Another study assigned subjects one of three explicit goals: ignore the faces, avoid using race to make their judgments, or intentionally use race to make their judgments (Payne et al., 2002). Results showed weapons bias in all conditions, but greater bias in the latter two conditions. There was no difference between the avoid-race and use-race conditions, despite the difference in intentions.

In a similar vein, Correll and colleagues (2002) have used a videogame simulation in which White and Black individuals were superimposed against a digital background, with either guns or harmless objects in their hands. Participants were instructed to identify whether the target was armed, and hit buttons marked "shoot" or "don't shoot" as quickly as possible. They found that when the target was not armed, participants mistakenly shot him more often if he was Black than if he was White. Moreover, this bias was found even among Black participants. Studies examining individual differences in these biases have found that both implicit and explicit measures of racial prejudice (Payne, 2001; 2005) and measures of perceptions of stereotypes in the culture (Correll et al., 2002) were associated with greater weapon bias. Does mere exposure to American cultural associations breed this kind of bias?

These studies suggest that one such association is between Blacks and criminality. Eberhardt and colleagues (2004) have shown that this conceptual connection works in both directions: priming Black faces makes people think of crime, whereas priming

crime makes people think of Black individuals. Though many people would disavow the first link, they might be less familiar with the second, and be less vigilant against its influence.

Highlighting the practical significance of these results, Correll and colleagues (2007) recently used the same methodology among samples of both police officers and civilians. When under no time constraint, police officers were faster, more accurate, and more balanced in their shooting decisions than civilian populations. When the response time window was reduced to force snap decisions, both civilians and police officers were slower to shoot an unarmed White man and faster to shoot an armed Black man. Yet police officers were unique in their ability to keep this automatic bias in response time from influencing their decisions to shoot, which remained less biased than the civilian sample. There were three correlates of the latency effect among police officers: the size of the community patrolled, its crime rate, and its proportion of Black civilians. Simply as a function of working within a racially diverse community, police officers might become more biased in their associations with criminality, even if not in their decisions to shoot. An encouraging finding is that practice with shooting simulations and officer firearms training decreased bias in shooting decisions (Correll et al., 2007; Plant & Peruche, 2005; Plant et al., 2005). This suggests that as new officers gain experience and training they begin to respond less like civilians and more like experts.

Implicit biases may nonetheless influence the decisions of even trained professionals in the criminal justice system. Using a randomly drawn sample of inmate records, Blair and colleagues (2004a) found that even though Blacks and Whites received equivalent sentences given equivalent criminal histories, those inmates with more Afrocentric features (regardless of actual race) received longer sentences. Subsequent studies showed that this kind of bias – in which targets with more Afrocentric features are assigned more stereotypical and negative traits – operates automatically. The bias occurs even under cognitive load, suggesting that it is efficient (Blair et al.,

2004b). When subjects were instructed not to use race in social judgment their implicit race bias but not Afrocentric feature bias decreased, suggesting that this bias is uncontrollable. Finally, even when given a prior task to rate faces based on Afrocentric features, subjects still could not control against the bias. This finding has stark implications. Because people tend to find the Afrocentric feature bias surprising and counterintuitive, they may be less successful at correcting for its effects than when correcting for more general implicit race bias.

The finding that Afrocentric features were associated with longer prison sentences is also striking because these studies examined actual prison sentences, not sentences suggested by research subjects in a mock trial exercise. This suggests that the defendants' features not only influenced research subjects, but also the judges who determined their sentences. The ecological relevance of these findings is highlighted even more dramatically in recent research on disparities in capital punishment. Eberhardt and colleagues (2006) found that defendants with more Afrocentric features were more likely to be sentenced to death, particularly when the victim of the crime was White.

These studies provide converging evidence that the influence of implicit bias can be detected in treatment of individuals in the criminal justice system. Criminal justice is an important arena for examining race bias because it has a long history of discriminatory treatment, and the consequences for both inmates and society at large can be severe. But findings of implicit bias are by no means limited to the criminal justice system. Several studies have found that implicit measures of bias are associated with discrimination in employment decisions.

Implicit bias and employment discrimination. A key motivation for developing implicit measures of bias was the observation that self-reported prejudice has steadily declined for decades, whereas evidence of discrimination persists (Schuhman et al., 1997; Sniderman & Carmines, 1997). The best evidence for persistent discrimination comes from field studies in which testers of different races are matched on relevant features and sent to do things

like rent apartments, buy cars, or apply for jobs (for a review, see Blank, Dabady, & Citro, 2004). For example, Bertrand and Mullainathan (2004) sent resumes to more than 1,300 help-wanted advertisements in Boston and Chicago. They manipulated whether the applicants were highly qualified or not, and whether the names on the resumes implied that the applicant was Black or White. They then measured the rate of callbacks, finding that White applicants received more calls than Black applicants. Specifically, White applicants were called back in 10.08% of cases, whereas Black applicants were called in only 6.79%, with the difference of 3.35% representing a 50% increased odds for White over Black applications. Moreover, qualifications mattered for White applicants but not Black applicants. Such findings, paired with very low rates of explicitly endorsed hiring discrimination, suggest that discrimination may be driven by implicit biases. However, these studies cannot directly test the link between implicit bias and hiring discrimination, a question that has been pursued in laboratory studies.

Ziegert and Hanges (2005) simulated hiring decisions among college participants. Cast in the role of rating job candidates, participants were put into either a “climate for bias” or a neutral climate. Climate was manipulated by the presence or absence of a memo from the “President” of the organization encouraging racial discrimination to preserve organizational stability. Implicit bias as measured by the IAT predicted discriminatory ratings of Black job candidates, but only when a climate for bias had been established. The social justice implications of implicit cognition are not limited to race. Other studies have taken different approaches to study the link between implicit gender bias and employment decisions. Rudman and Glick (2001) had participants evaluate female job applicants and complete a gender stereotype IAT measuring implicit associations pairing men and women with either communality or agency. Rather than directly influencing hireability ratings, implicit stereotypes of women as communal and men as agentic reduced perceived social skills of agentic women who were applying for “feminized”

jobs. This bias in turn mediated hiring discrimination for feminized jobs. In short, implicit biases appear to subtly influence employment decisions. Demonstrations of implicit bias in criminal justice and employment domains have been joined lately by studies suggesting that implicit bias may play a role in healthcare disparities as well.

Implicit bias and healthcare. The field of medicine strives to be fair and sympathetic to all its patients. Yet it seems possible that implicit social cognition will motivate frustration with “difficult” patients. One such stigmatized group is chronic drug users. In one recent study of healthcare, von Hippel and colleagues (2008) asked drug and alcohol nurses to report their explicit attitudes toward injecting drug users, their job-related stress, and intent to change jobs. Nurses also completed a Single Category IAT measuring implicit negative attitudes toward drug users. Implicit bias – but not explicit prejudice – mediated the relationship between job stress and intentions to change jobs.

One other study has documented effects of implicit bias on healthcare decisions. Green and colleagues (2007) had emergency and internal medicine residents read a vignette describing a Black or White patient’s symptoms, which included chest pains and a test result suggestive of a heart attack. Residents judged the likelihood of coronary artery disease, whether they would treat it with thrombolysis (a method using drugs to dissolve blood clots), and if so how strongly they felt. Residents also completed three IATs: an evaluative race IAT, an IAT measuring general stereotypes about Blacks being uncooperative, and an IAT measuring specific stereotypes about Blacks being medically uncooperative.

The first finding of interest was that all the IAT measures revealed implicit bias against Black patients among physicians, and these implicit biases were stronger than self-reported explicit biases toward Black patients. Secondly, and not accounting for implicit bias, physicians were more likely to diagnose Black patients than White patients with coronary artery disease, yet were equally likely to give thrombolysis as a treatment,

suggesting a disparity in treatment choice. Most important, implicit bias (as either the race IAT or a composite of the three IATs) predicted the decision to give thrombolysis even after accounting for explicit prejudice, demographic variables, and belief in treatment efficacy. Most interesting was the interaction of implicit bias and patient race. Doctors who were low in implicit bias gave thrombolysis more often to Black than to White patients; yet those who were high in implicit bias gave it to Black and White patients about equally. Though Black patients were diagnosed more often with heart disease, doctors high in implicit bias did not recommend potentially life-saving interventions that would track this demographic pattern.

Still, the specific pattern of findings in this study also raises new questions. As noted, doctors high in implicit race bias treated White and Black patients nearly equivalently. In contrast, doctors lowest in implicit bias were more likely to prescribe thrombolytic treatment to Black patients than Whites. So even though implicit bias accounted for treatment decision above and beyond other variables, the meaning of the correlation between IAT scores and treatment decisions remains ambiguous. More research is needed to understand whether the kinds of bias detected in this study contribute to explaining well-documented disparities in healthcare between Black and White patients.

Summary of implicit race bias

Using many different ways of assessing implicit bias and many ways of assessing behavior, these studies have demonstrated implicit bias that was associated with behaviors in the important domains of criminal justice, employment, and healthcare. Taken together, the evidence is much stronger now than even a few years ago. Implicit bias not only predicts subtle and ambiguous behaviors, but also meaningful behaviors with important consequences. As the field of implicit bias continues to quickly advance, its relevance for social justice is, in our view, only becoming clearer. Implicit cognition drives perceptions of Black individuals as both more criminal and less human.

Implicit biases create scenarios in which individuals may systematically discriminate against minorities without or counter to intent. Nonetheless, ordinary conceptions of justice require that a person must intend an act to be held responsible for it. Yet victims of such discrimination have a justified expectation to be treated based on their conduct as an individual rather than by their group membership. This dilemma sets the stage for disagreements about which behaviors are and are not morally acceptable. That being said, prejudice is not the only way that implicit social cognition influences social justice. As we describe in the next sections, judgments about who deserves moral treatment, and even about what acts are moral and just themselves, are products of implicit social cognition.

Implicit Cognition, Cultural Values, and Moral Judgment

In the previous section, we used intergroup conflict to show how implicit social cognition can engender behaviors that many would deem unethical. In the next section we review evidence that implicit processes can influence the values people adopt and the moral judgments they make. Implicit cognition not only leads to potentially unethical behavior, but also influences what we deem to be unethical.

Nationalist Ideologies

It has long been argued that there are two kinds of morality: one for our fellow group members, and one for everyone else (Cohen, Montoya, & Insko, 2006; Le Bon, 1896). According to anthropologist Margaret Mead, most pre-industrial societies do not recognize members of other tribal groups as fully human, a sentiment echoed in the findings on race and de-humanization described above. Mead claimed that “most primitive tribes feel that if you run across one of these subhumans from a rival group in the forest, the most appropriate thing to do is bludgeon him to death” (as cited in Bloom, 1997, p. 74). People have greater moral regard for members of their own groups, even in modern

industrialized societies (Cohen, Montoya, & Insko, 2006). However, people belong to many overlapping groups, and who is considered part of the ingroup can vary from one situation to another. There is evidence that in some cases, people's implicit judgments of who counts as an ingroup member may differ from their explicitly considered judgments. For example, who counts as an American?

Devos and Banaji (2005) found that when asked to explicitly define American identity, American participants rated emotional attachment to the nation and civic values as most important, and did not refer to ethnicity. Yet on an implicit test, Asians were less easily associated with American than Whites and Blacks, and Blacks slightly less so than Whites. Another study used faces of famous Asian Americans (e.g., Connie Chung) and White Europeans (e.g., Hugh Grant). Despite explicitly recognizing that the Asians were American and the Whites were European, participants still showed the White/American bias on an IAT measure. In a follow-up study, most American participants were ironically more likely to associate American with White than with Native American on the IAT (Devos, Nosek, & Banaji, unpublished). This research suggests that the very concept of what it means to be an American is laced with (and might encourage) implicit ethnocentrism.

This ethnocentrism has implications for foreign policy judgments. Uhlmann and colleagues (2009) had participants evaluate the moral justifiability of "collateral damage" in the Iraq war, and manipulated whether the innocent victims killed were Iraqis or Americans. They primed participants with either patriotic words or multi-cultural words in a sentence-unscrambling task. The question was whether making patriotism (as opposed to multiculturalism) accessible increased tolerance for Iraqi casualties. As expected, participants primed with patriotic words were more accepting of unintended casualties when they were Iraqi but not when they were American. Independent of this priming effect, conservatives were more tolerant of "collateral damage," especially when the victims were Iraqi. These findings suggest that, independent

of pre-existing ideologies, simply activating patriotic concepts may not only increase devotion to one's own country, but also increase ethnocentric leanings.

Nationalistic symbols have been shown to implicitly influence a range of outcomes. Ferguson and Hassin (2007; see also Ferguson et al., 2008) subliminally exposed American participants to images of the American flag. This led participants to rate the materialistic attributes of potential jobs as more important than other attributes. In another study, participants primed with an American flag in the experiment room completed word fragments more often with aggression- and war-related words, compared to when no flag was present. Finally, subliminally priming the American flag led to more hostile behavior in response to a mild computer-based provocation (though self-reported affect levels suggested no awareness of this). Across these studies, effects emerged only for those who were high in exposure to national news, a prominent source of cultural associations.

This illuminating set of studies reveals how something as simple as the American flag can encourage a range of feelings, perceptions, and behaviors that are not semantically related to flags in any obvious way. The link may instead be the larger set of values and ideology associated with the flag's symbolic meaning. Most striking of all, political orientation had no moderating effect in any of these studies. This suggests that patriotic concepts and symbols can activate aggressive nationalistic tendencies even among individuals who do not normally endorse these values.

In some cases, people act on activated values even when they contradict their own self-interest, as shown in studies of system justification. System justification theory claims that individuals at all levels in social hierarchies tend to be motivated to justify and defend the status quo (Blasi & Jost, 2006). This trend is often manifested historically when a society's current wrongdoings are rationalized away by its citizens (only to be recognized as wrong in retrospect, once the status quo has changed; see Hanson & Hanson, 2006). There is a related and curious tendency for members

of low-status groups to justify the very systems that prevent their advancement. Consider outgroup favoritism, or the tendency to prefer another group to one's own. In one study, students at high-status (Stanford) and low-status (San Jose State) colleges were administered IATs measuring attitudes toward each school, college stereotyping, and self-esteem (Jost et al., 2002). Though both sets of students displayed ingroup favoritism on average, a higher proportion of students at the low-status college showed outgroup favoritism. Implicit stereotyping of Stanford as academic and San Jose State as athletic was also positively correlated with outgroup favoritism among San Jose students. Finally, San Jose students with the lowest implicit self-esteem showed the greatest outgroup favoritism. This pattern of stereotyping and outgroup favoritism among lower status groups serves to support and sustain status differences.

System justification can also be seen in the "paradox of the free market": the faith in the legitimacy of the free market system among the poor, despite the growing inequality between the rich and poor. Support for fair market ideology predicted minimization of the ethical importance of the Enron scandal and its potential ties to the Bush-Cheney administration. The correlation was significant even when controlling for political conservatism. Jost and colleagues (2003) primed business students with information about companies that had small or large profits or losses. When fictional company names were used, students judged profitable companies to be more ethical and losing companies to be less ethical. When actual company names were used, any large deviation from the status quo (either as profit or loss) was seen as less ethical than smaller deviations. Finally, in a study among Hungarians, Jost et al. induced a threat either to the current economic status quo (capitalism) or the recent status quo (communism). Those under any kind of system threat increased their support for the free market system. These studies suggest that system justification motivates the tendency to see current economic systems as fair and just, overlooking ethical failures on the part of those systems.

Status differences are maintained also by complementary stereotypes, in which the negative consequences of membership in a low status group are perceived as offset by positive aspects (Kay et al., 2007). These stereotypes, such as poor-but-happy and rich-but-dishonest, have been found to persist on both the explicit and implicit levels (Kay et al., 2007). Kay and Jost (2003) primed participants with descriptions of individuals who fit complementary stereotypes (poor but happy and rich but unhappy; poor but honest and rich but dishonest) and measured explicit levels of system justification. Compared to when non-complementary stereotypes were presented (poor but unhappy, rich but happy), participants thought the general social system was more just. Moreover, on a lexical decision task the participants who had been primed with non-complementary stereotypes were faster at identifying justice-related as compared to neutral words, suggesting the activation of a justice motive. When complementary stereotypes are used in the context of economic inequality, people's sense of injustice dissipates. The fact that these patterns are found among both liberals and conservatives, and among low status as well as high status groups, suggests that system justification can lead to a defense of the ethical as whatever is in line with the status quo, even if this contradicts other explicit values and interests.

Where, then, do these values come from? The finding that exposure to news media moderated the effects in the Ferguson et al. (2008) studies suggests news media as one important way that these cultural values are transmitted. Research also suggests that implicitly activated values can be influenced by the present *and* the past. Uhlmann and colleagues (2008) have suggested that Americans are "implicit puritans." That is, their implicit associations are rooted in a web of values based on individual merit and traditional gender roles. In one study, participants primed with words related to spiritual salvation on a sentence-unscrambling task were more likely to persist in a subsequent anagram task. This effect only held for American, as opposed to Canadian, Italian, and German participants. This finding suggests that for Americans, there is an

implicit cognitive association between God and hard work.

In another study, Asian American students primed with work concepts in a sentence-unscrambling task endorsed traditionalist sex values, but only when their American (v. Asian) identities had been made salient. In a related study, participants read about a person who either supported or rejected a traditional American value, and were then given ambiguous information about her sexual practices. On a memory test, Americans primed with a person that rejected American work values falsely remembered that person as violating traditional sex values. It should be emphasized that all of these priming effects emerged even among liberals and the non-religious, suggesting a lingering influence of cultural associations even if not explicitly endorsed.

Another vestige of implicit Puritanism is moral absolutism, or the tendency to see the world as a Manichean division of good and evil. As we already saw in the section on intergroup bias, the tendency to take one's own perspective as the objective truth, either factually or morally, leads to rifts between social groups (Pronin, Gilovich, & Ross, 2004). Values are those ends that we hold most dear, yet the research just described suggests that we perceive and act based on values that we might not consider our own. In the next section we review research that questions whether people have a clear idea of exactly which moral values ground their judgments.

Implicit Cognition and the Feeling of Moral Clarity

An emerging consensus claims that people's moral judgments are driven by automatic, intuitive processes. Haidt's (2001) Social Intuitionist Model was the first to seriously incorporate the lessons of implicit social cognition. Haidt suggests that the immediate response to a moral transgression is a moral intuition, which is "the sudden appearance in consciousness, or at the fringes of consciousness, of an evaluative feeling about the character or actions of a person, without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion" (Haidt &

Bjorklund, 2008, p. 188). Moral intuition – which can but need not involve emotion – is the immediate precursor of moral judgment, which is "the conscious experience of blame/praise, including belief in the rightness or wrongness of an act" (Haidt & Bjorklund, 2008, p. 188). According to this model, moral reasoning enters primarily *after* the moral judgment has been made, finding post hoc reasons to support the original intuition. Although deliberate reasoning can play a causal role in reaching moral judgments, it is a rather subordinate role in producing and refining moral judgment.

The strongest support for this model comes from research on moral dumbfounding (Haidt & Bjorklund, 2008). In this paradigm, participants are asked to evaluate moral transgressions such as incest, cleaning a toilet with the national flag, and eating one's own dog. Participants are then asked to justify their moral responses, but the experimenter systematically shifts the scenario to counteract all of the possible practical justifications (e.g., incestuous siblings used birth control; it was only a one-time occurrence; neither was psychologically harmed by the experience). Participants typically revert to the position that although they cannot explain why they feel the way they do, the act is nevertheless simply wrong.

This dissociation between moral judgment and justification has been replicated and expanded upon by Hauser's research into innate moral grammar. Hauser's model emerged in response to an emotion-based reading of Haidt's theory: if emotion drives moral judgment, what drives the emotional response? Hauser (2006) has suggested that all humans have a store of operative knowledge about morality, a set of unconscious principles that lead to automatic moral judgments in the face of transgressions. Many of these principles are inaccessible to conscious awareness, and operate over appraisals of intent and cause to create judgments of rightness and wrongness. On this view, emotions and conscious reasoning are the byproducts of this automatic appraisal process.

To test these claims, Cushman and colleagues (2006) explored the conscious accessibility of three moral principles: the action principle (harmful

actions are worse than harmful omissions), the intention principle (harm as a means to an end is worse than harm as foreseen – but unintended – side effect), and the contact principle (using physical contact to harm is worse than causing harm without contact). Participants were asked to make moral decisions in response to scenarios that manipulated these three principles, and then to justify those decisions. Judges then coded how well the justifications could account for the causal effects of the three principles on moral judgments. Although moral judgments aligned with the three principles, justifications were more varied. Most participants gave sufficient justification for judgments based on the action principle, suggesting they had conscious awareness of it. More than half of participants were able to justify their decisions based in the contact principle. Finally, less than a third of participants sufficiently justified their judgments based on the intention principle. The breakdown in justification for the intention principle has been replicated in a cross-cultural sample as well (Hauser et al., 2007).

These studies suggest that people's justifications about the reasons they made a particular moral judgment sometimes track the actual causal impact of the underlying principles, but they do not necessarily do so. Together, the studies by Haidt, Hauser, and Cushman show that when people attempt to explain the basis for their moral judgments, they sometimes can give a coherent explanation, but just as often they seem to confabulate reasons that bear little relationship to the underlying causes of their judgments.

Some have criticized these models for their neglect of controlled cognition. For instance, Greene (2008) has argued that at least two processes are involved in moral judgment: automatic emotional responses and controlled cognitive deliberation. To test this, he subjected participants to a cognitive load manipulation while they read a series of dilemmas which pitted strong emotion against utilitarian analysis (e.g., would you smother your baby in order to prevent you and your entire family from being caught and killed by a death squad?) Cognitive load made participants take longer to give a utilitarian response – but not the

emotional response – to these moral dilemmas, suggesting interference with utilitarian reasoning but not emotion-based judgment.

Yet the exact roles of automatic and controlled processes in moral judgment are still a matter of debate. For example, Epley and Caruso (2004) speculated that because moral judgments are based in egocentric and automatic evaluations of the environment as good or bad, these judgments will tend to be self-serving. Subsequent moral disagreements between groups will latch onto post hoc justifications which may have no bearing on the real issues at hand. On the other hand, DeSteno and colleagues (2008) recently investigated the processes underlying moral hypocrisy, the tendency to say that a moral violation is more permissible when committed by oneself than by others. In contrast to the perspective that automatic moral judgments are always self-serving, they found that placing participants under cognitive load erased tendencies toward moral hypocrisy. This suggests that attention-demanding processes of motivated reasoning may underlie this selectivity. Although research on implicit and explicit aspects of moral judgment is still new, initial evidence suggests that both automatic intuitive reactions and deliberate rationalizations may both contribute to moral judgments that seem from an egocentric perspective to be obviously correct, but that seem to others to be biased or arbitrary.

The fact that implicit cognition drives at least some aspects of moral judgment raises the question of whether people should trust their moral intuitions. Gigerenzer (2008) argues that “moral heuristics,” including the moral intuitions in Haidt's model, can be made more conscious and avoided if necessary. Yet he argues this won't often be necessary, because the heuristics adaptively simplify the moral sphere. Sunstein (2005), by contrast, has emphasized how these heuristics can be taken out of context, reified as moral principles, and misapplied to improper situations. For example, many people have an intuitive disgust reaction toward marginalized social groups that translates into moral disapproval. If consciously recognized, many people would disavow this emotional

heuristic as a justification for moral disapproval (Schnall et al., 2008). The key point for our purpose is that people's intuitive moral reactions may sometimes differ starkly from their explicitly endorsed moral principles.

One recent pair of studies attests to this idea. Inbar and colleagues (in press) investigated intuitive moral disapproval of homosexuals. In their first study, participants read about a film director who encouraged either gay or straight kissing. Explicitly, most people indicated that nothing was morally wrong with the kissing. Yet participants viewed the director's action as more intentional when he encouraged gay kissing than when he encouraged straight kissing, and this effect was strongest for those high in disgust sensitivity. Based on Knobe's finding (2006) that people are more likely to say an action is intentional if they conceive of it as morally wrong, this indirect approach suggests that people might be making an implicit moral wrongness judgment that contradicts their explicitly stated values. In a second study using an IAT which paired gay/straight with pleasant/unpleasant, people were faster to associate gay with unpleasant, especially when they were high in disgust sensitivity. In short, many people show traces of implicit moral disapproval of homosexuals, against their explicitly stated values.

How much, then, should people trust their moral intuitions? Singer (2005), Greene (2008), and others argue that we should not trust our moral intuitions until we reflectively validate them against some external criterion. But external criteria are hard to come by, especially concerning moral judgments (Pizarro & Uhlmann, 2005). To the extent that people are naïve realists who simply "know" that they are right, this validation process may be difficult or impossible (Pronin, Gilovich, & Ross, 2004). One can think of plenty of post hoc reasons to support those intuitions, reasons which may subjectively feel as valid as any other.

There are striking similarities between this research on moral judgment and research on implicit prejudice and other social attitudes. In both cases, implicit processes sometimes produce judgments that are difficult to justify and explain.

Implicitly measured moral sentiments tend to show less tolerance and more hierarchy. Implicitly measured sentiments also tend to reflect historically prominent attitudes and values that may no longer be explicitly endorsed. Theories of moral intuitions have recognized that automatic and controlled processes interact to produce moral judgments, but the specific roles for these processes have rarely been specified. It could be the case that implicit moral cognition works the same way as implicit racial cognition, in which case lessons about the self-regulation of prejudiced impulses might prove fruitful in understanding the role of controlled deliberation in moral judgment (Fine, 2006; Kennett & Fine, 2009). Research in moral cognition might be profitably expanded by integrating these insights with dual process theories that have been developed in other domains of social cognition.

A recently developed set of quantitative process models offers promise in this regard. Multinomial process models allow researchers to empirically test hypotheses about the relative dominance of automatic and controlled influences on judgments (Bishara & Payne, in press; Conrey et al., 2005; Payne, 2001, 2005; Payne & Bishara, 2008; Sherman et al., 2008). For example, in automaticity-dominating models, whenever an automatic response is activated, it alone determines responses. Controlled efforts (for example, responding based on deliberate reasoning) would come into play only in the absence of an automatic moral reaction. In contrast, in control-dominating models, whenever deliberate reasoning is engaged, it alone determines responses. But when deliberate reasoning fails, automatic moral intuitions drive the response. These models differ in which process – automatic moral intuitions versus deliberate moral reasoning – dominate responding when they conflict. Such models could allow empirical tests of the relative strength of automatic and controlled aspects of moral judgment (e.g., Conway & Gawronski, 2009).

Answering the question of which process dominates the other, and under what conditions, is relevant for determining how prominent a role moral reasoning plays in moral judgment. According to some theories such as Kohlberg's

(1971) or Turiel's (1983), moral reasoning might be expected to be the dominant process. But according to other theories such as Haidt's (2001) or Hauser's (2006), automatic intuitions can be expected to play the dominant role. The more prominent the role of automatic intuitions, the more reason there is to expect that moral intuitions will commonly conflict with traditional rationalist principles and reflectively endorsed values.

***Empirical Research and Normative Implications:
The Role of Is and Ought***

In the preceding pages we have reviewed research demonstrating the interplay between implicit social cognition and matters of social justice and morality. We showed how implicit race bias sometimes leads people to act in ways that run against their explicitly considered values of tolerance. We showed how implicit thought processes can exclude certain groups from one's own moral circle, and operate to maintain rigid nationalistic, ethnocentric, and hierarchy-based boundaries. Finally, we described research showing that implicit cognition shapes what people consider moral and just in the first place. In each of these cases, the empirical research identifies ordinary situations that create tensions between typical conceptions of moral judgment and actual behaviors driven by implicit processes. What are we to make of these tensions?

These tensions suggest that people often have less clarity over their mental lives than they would have themselves (and others) believe. On the one hand, this has clear epistemic implications for knowledge of oneself and others. Yet whether this fact has moral implications is decidedly less clear. Does wrongful certainty lead to certain wrongs? It has been said many times that you can't get an 'ought' from an 'is'. As David Hume argued, empirical facts about how the world is do not determine how it ought to be. Considering implicit social cognition research, none of the findings themselves can dictate whether any aspect of behavior is morally right or wrong. Yet even though the empirical findings do not themselves establish any moral standards, the findings have direct

relevance for what societies and individuals believe is morally acceptable. Knowing that implicit cognition can alter who we judge to have moral status, or what we judge to be a moral issue at all, can substantively inform moral dialogue. Knowing how implicit cognition can cause our ethicality to corrode can also help us engage better moral self-regulation in pursuit of our ideals. Normative moral standards are value judgments made by communities (Gibbard, 1990). The role of empirical facts is to ground those judgments so that communities are engaged in informed (rather than uninformed) conversations.

Implicit social cognition research is particularly well-suited to establish facts about the role of morally relevant mental states, especially intent, conscious awareness, and controllability. Lay moral intuitions and formal legal codes distinguish between intentional versus unintentional acts, between conscious versus unconscious acts, and between controllable versus uncontrollable acts. If racial discrimination, for example, were found to be always intentional, conscious, and controllable then presumably it would always be considered wrong. If it were found to sometimes be unconscious or unintentional, then the possibility arises that someone may discriminate without being considered morally responsible. Wherever moral or legal judgments hinge on questions of intent, consciousness, and controllability, implicit social cognition research becomes relevant.

These considerations affect policy choices as well. Kang and Banaji (2006), for instance, have argued that the continued existence of implicit race bias provides a new and different rationale for affirmative action policies. Rather than viewing affirmative action as a remedy for historical injustice, research showing continued subtle forms of discrimination could be used to justify affirmative action as a remedy for present and on-going discrimination. In this case, implicit social cognition research has been used to argue for policies aimed at reducing discrimination. Yet there is no necessary connection between these findings and the values or goals for which they are used. It is easy to imagine someone discriminating and then

using the findings of implicit bias to argue that he or she is not responsible. Here again, the empirical research does not itself dictate the proper response, but it provides the scaffolds on which arguments are built.

To date, most of the discussion about the moral and legal implications of implicit social cognition has taken place in the pages of academic books and journals (Bargh, 1999; Fiske, 2005; Kelly & Roedder, 2008). But as behavioral science findings become more widely known, the popular understanding of these findings is likely to have broader impact. Similar trends can be seen for forensic science. Popular television programs such as the CBS series *CSI: Crime Scene Investigation* have had a noticeable impact on the kinds of evidence that juries expect. Jury members familiar with the forensic techniques used in the show have begun demanding higher and higher levels of forensic evidence such as DNA analysis before rendering a guilty verdict (Thomas, 2006). Although implicit social cognition research has not reached the prominence of this television series, it has gained relatively high notoriety for a behavioral science. Thanks in part to the popularity of the *Project Implicit* website, millions of individuals have completed some form of implicit test. Implicit biases have been discussed many times in the pages of the *New York Times*, the *Wall Street Journal*, *Time* and *Newsweek*. As information about implicit social cognition becomes popular knowledge, we can expect to see effects on the public's judgments and decisions.

We conclude by describing a study that takes a look at what kind of an impact such knowledge might have. Cameron, Payne, & Knobe (2008) asked subjects to read about cases of racial discrimination. The same discriminatory behaviors happened in all conditions, but the awareness and controllability of the acts were manipulated. In a control condition, no information about the actor's mental states was provided. In a second condition, the actor was described as being consciously aware of his dislike for African Americans, but he rejected that feeling and made an effort to treat people equally. Still, he ended up discriminating

unintentionally because prejudice influenced his actions in ways he couldn't fully control. This description reflected theories of implicit bias that emphasize automaticity rather than unconsciousness. Finally, in a third condition subjects read about an actor who had a completely unconscious dislike for African Americans. He discriminated in the same way as in the other conditions because he had no awareness of it. This condition reflected theories of implicit bias that emphasize unconscious attitudes.

Participants made judgments of how morally blameworthy the discrimination was. When the discrimination was conscious but uncontrollable, it was judged to be only slightly less blameworthy than when no information about mental states was given. But when the discrimination was described as stemming from unconscious bias, it was judged much less blameworthy. Subjects gave a great deal of weight to the conscious awareness of the bias leading to discrimination, much more than they gave to its controllability. In sum, when implicit biases are represented as being unconscious, the discrimination that results from them is blamed significantly less. Ironically, the strategy of raising consciousness of one's biases (Banaji, Bazerman, & Chugh, 2003) may expose a person to more blame than if the biases remain outside of awareness.

This study takes a first step toward understanding the likely impact of implicit social cognition research by showing that the particular theory used to explain the findings can have a strong impact on lay judgments. The stakes are high, because the image of the human mind developed in this research is likely to influence social, legal, and political decisions. The research reviewed here highlights the importance of making careful distinctions between specific aspects of automatic thinking, such as distinguishing awareness, intention, and controllability. People do appear to make these distinctions in their lay moral judgments. In our view, the key lesson for researchers is that implicit social cognition research makes a difference in collective conversations about social justice. The stronger the empirical base, the better informed is the conversation.

References

- Arbuckle, N., & Payne, B.K. (2008). Implicit intergroup emotions. Unpublished manuscript.
- Arkes, H., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or Would Jesse Jackson fail the Implicit Association Test? *Psychological Inquiry*, *15*, 257-278.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*, 193-209.
- Banaji, M.R., Bazerman, M.H., & Chugh, D. (2003). How (un)ethical are you? *Harvard Business Review*, *81*, 56-64.
- Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes & Tetlock. *Psychological Inquiry*, *15*, 279-289.
- Bargh, J.A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 361-382). New York: Guilford Press.
- Baumeister, R. F., Vohs, K. D., DeWall, C. N., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, *11*, 167-203.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, *94*, 991-1013.
- Bishara, A.J., & Payne, B.K. (2009). Multinomial process tree models of control and automaticity in weapon misidentification. *Journal of Experimental Social Psychology*, *45*, 524-534.
- Blair, I.V., Judd, C.M., & Chapleau, K.M. (2004a). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*, 674-679.
- Blair, I.V., Judd, C.M., & Fallman, J.L. (2004b). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, *87*, 763-778.
- Blank, L.M., Dabady, M., & Citro, C.F. (2004). *Measuring Racial Discrimination*. National Academies Press.
- Blasi, G., & Jost, J.T. (2006). System justification theory and research: Implications for law, legal advocacy, and social justice. *California Law Review*, *94*, 1119-1168.
- Bloom, H. (1997). *The Lucifer principle: A scientific expedition into the forces of history*. New York: Atlantic Monthly Press.
- Cameron, C. D., Payne, B. K., & Knobe, J. (2008). Do theories of implicit race bias influence moral judgments? Unpublished manuscript.
- Chen, M. & Bargh, J.A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, *33*, 541-560.
- Cohen, T., Montoya, R., & Insko, C. (2006). Group morality and intergroup relations: Cross-cultural and experimental evidence. *Personality and Social Psychology Bulletin*, *32*, 1559-1572.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469-487.
- Conway, P.J., & Gawronski, B. (2009, February). *Deontological vs. utilitarian inclinations in moral decision making: A process dissociation approach*. Poster presented at the 10th Annual Meeting of the Society for Personality and Social Psychology (SPSP), Tampa, FL, USA.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*, 1314-1329.

- Correll, J., Park, B., Judd, C.M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality & Social Psychology*, *92*, 1006-1023.
- Cushman, F.A., Young, L., & Hauser, M.D. (2006). The role of reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science*, *17*, 1082- 1089.
- Devos, T., & Banaji, M.R. (2005). American=White? *Journal of Personality and Social Psychology*, *88*, 447-466.
- Devos, T., Nosek, B. A., & Banaji, M. R. (2007). Aliens in their own land? Implicit and explicit ascriptions of national identity to Native Americans and White Americans. Unpublished manuscript.
- Dovidio, J.F., Kawakami, K., & Gaertner, S.L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62-68.
- 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.
- Eberhardt, J.L., Goff, P.A., Purdie, V.J., & Davies, P.G. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, *87*, 876-893.
- Epley, N., & Caruso, E. (2004). Egocentric ethics. *Social Justice Research*, *17*, 171-187.
- Fazio, R.H., Jackson, J.R., Dunton, B.C., & Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes. *Journal of Personality and Social Psychology*, *69*, 1013-1027.
- Ferguson, M.J., Carter, T.J., & Hassin, R.R. (2008). On the automaticity of American nationalism. In J.T. Jost, A.C. Kay, & H. Thorisdottir (Eds.), *Social and psychological bases of ideology and system justification*. New York: Oxford University Press.
- Ferguson, M.J., & Hassin, R.R. (2007). On the automatic association between America and aggression for news watchers. *Personality and Social Psychology Bulletin*, *33*, 1632-1647.
- Fine, C. (2006). Is the emotional dog wagging its rational tail, or chasing it? *Philosophical Explorations*, *9*, 83-98.
- Fiske, S.T. (2005). What's in a category?: Responsibility, intent, and the avoidability of bias against outgroups. In A. Miller (Ed.), *The social psychology of good and evil* (pp. 127-140). New York: Guilford Press.
- Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass*, *2*, 1002-1023.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Goff, P.A., Eberhardt, J.L., Williams, M.J., & Jackson, M.C. (2008). Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences. *Journal of Personality and Social Psychology*, *94*, 292-306.
- Green, A.R., Carney, D.R., Pallin, D.J., Ngo, L.H., Raymond, K.L., Iezzoni, L.I., & Banaji, M.R. (2007). Implicit bias among physicians and its prediction of thrombolytic decisions for Black and White patients. *Journal of General Internal Medicine*, *22*, 1231-1238.
- Greene, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 3. The neuroscience of morality: Emotion, disease, and development* (pp. 35-79). Cambridge, MA: MIT Press.
- Greenwald, A.G., McGhee, D.E., & Schwarz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464-1480.
- Gigerenzer, G. (2008). Moral intuition = Fast and frugal heuristics?. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 2. The cognitive science of*

morality: Intuition and diversity (pp. 1-26). Cambridge, MA: MIT Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814-834.

Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 2. The cognitive science of morality: Intuition and diversity* (pp. 181-218). Cambridge, MA: MIT Press.

Hanson, J., & Hanson, K. (2006). The blame frame: Justifying (racial) injustice in America. *Harvard Civil Rights-Civil Liberties Law Review*, *41*, 414-480.

Hauser, M. (2006). *Moral minds*. New York: Harper-Collins Publishers.

Hauser, M., Cushman, F., Young, L., Jin, R.K., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind and Language*, *22*, 1-21.

Herszenhorn, D. M. (1997). Punitive actions are advised in discrimination at Denny's. New York Times online, August 15, 1997.

Inbar, Y., Pizarro, D.A., Knobe, J., & Bloom, P. (2009). Disgust sensitivity predicts intuitive disapproval of gays. *Emotion*, *9*, 435-439.

Jost, J.T., Blount, S., Pfeffer, J., & Hunyady, Gy. (2003). Fair market ideology: Its cognitive-motivational underpinnings. *Research in Organizational Behavior*, *25*, 53-91.

Kang, J. (2005). Trojan horses of race. *Harvard Law Review*, *118*, 1491-1593.

Kang, J., & Banaji, M.R. (2006). Fair measures: A behavioral realist revision of 'affirmative action.' *California Law Review*, *94*, 1063-1118.

Kay, A.C., & Jost, J.T. (2003). Complementary justice: Effects of "poor but happy" and "poor but honest" stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology*, *85*, 823-837.

Kay, A. C., Jost, J.T., Mandisodza, A.N., Sherman, S.J., Petrocelli, J.V., & Johnson, A.L. (2007). Panglossian ideology in the service of system justification: How complementary stereotypes help us to rationalize inequality. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 39, pp. 305-358). San Diego, CA: Elsevier.

Kelly, D., & Roedder, E. (2008). Racial cognition and the ethics of implicit bias. *Philosophy Compass*, *3*, 522-540.

Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin*, *34*, 833-848.

Kennett, J., & Fine, C. (2009). Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology. *Ethical Theory and Moral Practice*, *12*, 77-96.

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*, *130*, 203-231.

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the story of moral development. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151-235). New York: Academic Press.

Krieger, L.H., & Fiske, S.T. (2006). Behavioral realism in employment discrimination law: Implicit bias and disparate treatment. *California Law Review*, *94*, 997-1062.

Lane, K.A., Kang, J., & Banaji, M.R. (2007). Implicit social cognition and law. *Annual Review of Law and Social Science*, *3*, 427-451.

Le Bon, G. (1896). *The crowd: A study of the popular mind*. New York: The Macmillan Company.

Leyens, J., Paladino, P.M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups.

Personality and Social Psychology Review, 4, 186-197.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measure of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435-442.

Mitchell, G., & Tetlock, P.E. (2007). Antidiscrimination law and the perils of mindreading. *Ohio State Law Journal*, 67, 1023-1122.

Norton, M.I., Vandello, J.A., & Darley, J. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, 87, 817-831.

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion*, 22, 553-594.

Payne, B.K. (2001). Prejudice and perception: The role of automatic and controlled perceptions in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181-192.

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive control modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, 89, 488-503.

Payne, B. K., & Bishara, A. J. (in press). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*.

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

Payne, B.K., Lambert, A.J., & Jacoby, L.L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology*, 38, 384-396.

Pizarro, D.A., & Uhlmann, E. (2005). Do normative standards advance our understanding of moral judgment? *Behavioral and Brain Sciences*, 28, 558-559.

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

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

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111, 781-799.

Rudman, L.A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57, 743-762.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145-172.

Salvatore, J., & Shelton, N. (2007). Cognitive costs of exposure to racial prejudice. *Psychological Science*, 18, 810-815.

Schnall, S., Haidt, J., Clore, G., & Jordan, A. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34, 1096-1109.

Schuhman, H., Steeh, C., Bobo, L., & Kyrsan, L. (1997). *Racial attitudes in America: Trends and interpretations*. Cambridge, MA: Harvard University Press.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, 115, 314-335.

Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics*, 9, 331-352.

Smith, A. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, 115, 236-271.

Sniderman, P. M., & Carmines, E. G. (1997). Reaching beyond race. *Political Science and Politics*, 30, 466-471.

Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613-629.

Sunstein, C. (2005). Moral heuristics. *Behavioral and Brain Sciences*, *28*, 531-573.

Thomas, A. P. (2006). The CSI Effect: Fact or Fiction, 115 Yale Law Journal Pocket Part 70, <http://www.thepocketpart.org/2006/02/thomas.html>.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.

Uhlmann, E.L., Brescoll, V.L., & Paluck, E.L. (2006). Are members of low status groups perceived as bad, or badly off? Egalitarian negative associations and automatic prejudice. *Journal of Experimental Social Psychology*, *42*, 491-499.

Uhlmann, E.L., Pizarro, D.A., Tannenbaum, D., & Ditto, P.H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, *4*, 476-491.

Uhlmann, E.L., Poehlman, T.A., & Bargh, J.A. (2008). American moral exceptionalism. In J.T Jost, A.C. Kay, & H. Thorisdottir, (Eds.), *Social and psychological bases of ideology and system justification*. New York: Oxford University Press.

Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, *44*, 1334-1338.

Von Hippel, W., Brener, L., & von Hippel, C. (2008). Implicit prejudice toward injecting drug users predicts intentions to change jobs among drug and alcohol nurses. *Psychological Science*, *19*, 7-11.

Word, C., Zanna, M., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, *10*, 109-120.

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.