

13 Mechanisms Underlying the Malleability of Implicit Prejudice and Stereotypes

The Role of Automaticity and Cognitive Control

Nilanjana Dasgupta
University of Massachusetts

A central theme in contemporary social psychology is that people's attitudes, beliefs, and behavior are often shaped by factors that lie outside their awareness and cannot be fully understood by intuitive methods such as self-reflection (Bargh, 1997; Greenwald & Banaji, 1995; Nisbett & Wilson, 1977). In the last 20 years, experimental social psychology has discovered an important window into mental life by discovering that attitudes and beliefs can be activated in memory without perceivers' awareness or intention. Once activated, these cognitions and evaluations are difficult to suppress or inhibit in the moment and create prepotent action tendencies that facilitate evaluation-consistent behavior, judgments, and decisions. These subtle reactions have been variously labeled *implicit*, *automatic*, *unconscious*, or *nonconscious* (Bargh, 1994; Greenwald & Banaji, 1995; Kihlstrom, 1990). However, because it is rare for any psychological judgment or behavior to meet all of these criteria at the same time—lack of awareness, intention, effort, and control—any judgment that meets at least one of these criteria has been given these labels.

Implicit attitudes and beliefs are typically seen as conceptually distinct from *explicit*, *controlled*, *self-reported*, or *conscious* responses. As these terms suggest, attitudes are considered explicit when perceivers are aware of their evaluations, able to endorse them as personally held opinions, and when they have the capacity to learn and change their attitudes volitionally by expending effort. Whereas explicit attitudes are measured by directly asking people to consider how they feel about a particular object or issue and then report their thoughts in a deliberate fashion, implicit attitudes are inferred indirectly from people's performance on tasks that, at face value, seem unrelated to attitude measurement. For example, the speed with which people associate certain stimuli on speeded reaction time tasks or their choice of words on word completion tasks is used to infer implicit attitudes.

Much of the research on implicit attitudes and their effects on social behavior has been conducted in the domain of intergroup relations, particularly around issues of prejudice and stereotyping. Research has gravitated in this direction for two theoretical reasons. First, the socially sensitive nature of intergroup thoughts and evaluations typically raises concerns that people's voluntary responses toward in- and outgroups may be distorted by self-presentation and impression management concerns. In other words, people may not always be willing to report socially sensitive attitudes honestly, especially if those attitudes deviate from social norms. Second, when self-reporting their attitudes people sometimes make a strong distinction between their own personal attitudes

and those circulating in the larger culture (“society at large is prejudiced against Group X, but I am not”). Yet, societal construals of particular groups may have been passively learned and incorporated into perceivers’ own mental representations without their knowledge. In other words, when asked, people may not have complete introspective access to their attitudes and thus may not be able to report them fully and accurately (Banaji & Greenwald, 1994; Greenwald & Banaji, 1995; Nisbett & Wilson, 1977).

IMPLICIT ATTITUDES ARE MALLEABLE

The empirical evidence that implicit attitudes are automatically activated without awareness, and that they have the capacity to drive judgments and behavior regardless of explicit intention and control, had, for a long time, led to the conclusion that these attitudes are relatively immutable. Early theories of implicit social cognition argued that implicit attitudes and beliefs are learned early in life and that they change slowly across time only after the accrual of new associations and a great deal of training (Bargh, 1999; Devine, 1989; Petty, Tormala, Brinol, & Jarvis, 2006; Wilson, Lindsey, & Schooler, 2000). In other words, the assumption was that conventional persuasion techniques that change explicit attitudes by relying on perceivers’ awareness of their attitudes, motivation to reconsider their stance, and willingness to expend effort to consider new information should leave implicit attitudes untouched.

As in the case of attitude change in general, prejudice reduction interventions that have been reported in the social science literature have typically assumed that conscious mental processes must be engaged for prejudicial attitudes to change. Specifically, the working assumption was that perceivers must: (a) be aware of their bias (Banaji, 2001; Dasgupta, 2004, *in press*); (b) be motivated to suppress negative thoughts (Macrae, Bodenhausen, Milne, & Jetten, 1994; Macrae, Bodenhausen, Milne, & Wheeler, 1996); (c) be motivated to change their responses toward outgroups because of personal values, feelings of guilt, compunction, or self-insight (Allport, 1954; Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, 1993; Monteith, Devine, & Zuwerink, 1993; Monteith, Zuwerink, & Devine, 1994; Myrdal, 1944); (d) exert effort to seek cognitive consistency between their general egalitarian values and attitudes toward specific groups (Gaertner & Dovidio, 1986; Katz & Hass, 1988; Katz, Wackenhut, & Hass, 1986; Roakeach, 1973); (e) develop and practice correction strategies to unlearn negative stereotypes (Gawronski, Deutsch, & Mbirikou, 2007; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000); and (f) be willing and motivated to engage in intergroup contact (Pettigrew & Tropp, 2006; Tropp & Bianchi, 2006; Tropp, Stout, Boatwain, Wright, & Pettigrew, 2006). Because changing intergroup attitudes was viewed as a self-conscious relearning process, the research just cited mostly focused on changing explicit attitudes. Until recently, few attempted to modify implicit forms of prejudice and stereotyping because these were seen as inescapable habits that are expressed despite attempts to bypass or ignore them (Bargh, 1999; Devine, 1989).

The advent of new data and new theories has cast doubt on the immutability of implicit attitudes and beliefs. The challenge has come from two sources. First, empirical evidence accumulating over the past 5 years has shown that implicit attitudes shift in response to various contextual and psychological factors (for reviews see Blair, 2002; Gawronski & Bodenhausen, 2006). Second, new theoretical models have begun to refine and modify the definition of implicit social cognition (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Gawronski & Bodenhausen, 2006). These theories identify the mechanisms underlying rapid reactions to in- and outgroups that occur under time pressure. These mechanisms, in turn, may help clarify why particular social contexts, internal psychological states, or individual differences evoke changes in implicit attitudes and beliefs.

COGNITIVE CONTROL INFLUENCES THE MALLEABILITY OF IMPLICIT ATTITUDES

Two theoretical models have focused on the role of cognitive control in shaping attitude expressions that are typically thought of as “implicit” (see Conrey et al., 2005, for a description of the quadruple process model; see Payne, 2001, 2005; Payne, Lambert, & Jacoby, 2002, for a description of the process dissociation procedure [PDP] as applied to implicit attitude tasks). Conrey, Payne, and their colleagues have argued that although implicit attitudes may be activated without awareness (as demonstrated by subliminal priming studies in which perceivers are unaware of seeing in- and outgroup images) and expressed under time pressure (as demonstrated by studies using speeded reaction time tasks that constrain the time allowed to respond), such responses do not rule out the role of controlled processes. In other words, attitudes measured by seemingly implicit tasks are not “process pure”; rather, they are guided by a blend of automatic and controlled processes. For example, in the case of reaction time tasks that use the speed with which people associate social groups with particular attributes to indirectly infer attitude strength, part of those speeded responses is driven by the activation of automatic associations but another part is determined by individuals’ ability to selectively attend to information that facilitates accurate responses and screen out unnecessary information that hinders accurate responses. By applying theories of cognitive control to the accumulated evidence demonstrating the malleability of implicit responses, one might ask this question: Do changes in cognitive control function as one mechanism responsible for the flexibility of implicit attitudes? Might particular social contexts or psychological states increase cognitive control and is this, in turn, responsible for the alleviation of implicit bias?

ACCESSIBILITY OF AUTOMATIC ASSOCIATIONS INFLUENCES THE MALLEABILITY OF IMPLICIT ATTITUDES

A different mechanism that may underlie the flexibility of implicit attitudes toward social groups has to do with changes in the automatic associations linking particular groups to particular attributes. Several theories have argued that automatic associations are learned through repeated exposure to certain group–attribute pairings in the larger society either via firsthand experience with group members who have certain characteristics or via mediated exposure from the mass media and information learned from peers and significant others (Gawronski & Bodenhausen, 2006; Smith & DeCoster, 2000; see also Rydell & McConnell, *in press*; Strack & Deutsch, 2004). For example, the associative-propositional evaluation model (APE model) proposes that once group–attribute associations are learned, they are likely to be activated automatically in the presence of a relevant target person irrespective of their perceived “truth value” (i.e., whether or not the perceiver considers these evaluations to be accurate). The implication here is that while repeated learning of counterstereotypic associations may change the original automatic associations, simply being told that the information one has learned about a target group is inaccurate (i.e., information about its truth value) should not change these associations (Gawronski & Bodenhausen, 2006).

Moreover, target groups may be associated with multiple attributes in memory. Which particular group–attribute association will become activated in the presence of a particular group member depends on the goodness of fit between preexisting mental associations and a particular set of external inputs. Thus, if a particular target group (e.g., Asians) is associated with two types of attributes (intelligent and nonathletic), which attribute will become activated in the presence of an Asian person will depend on the characteristics of that particular individual, the context in which he or she is encountered, and the goodness of fit between the external situation and the associations in memory. Seeing an Asian individual in a classroom is more likely to activate the “intelligent” association and influence subsequent judgments consistent with intelligence, whereas seeing the same individual on a soccer field is more likely to activate the “nonathletic” association and lead to different sorts of judgments. In other words, shifts in implicit attitudes about a particular group may be driven

by the activation of different types of automatic associations that already exist in memory. In this example, encountering a particular person in a particular situation is the trigger that activates one of the underlying associations, making it rise to threshold and get expressed in a judgment or social behavior (see also Smith & DeCoster, 2000).

These new theories (QUAD, APE, and PDP models) offer refined descriptions of implicit social cognition (attitudes, beliefs, knowledge). They suggest that implicit responses are driven by a mixture of automatic associations rendered accessible in the moment and executive control driven by internal states (e.g., motivations, goals, emotions, individual differences, etc.). Moreover, they assume that these two processes work independently to influence social behavior.

GOALS OF THIS CHAPTER

This chapter revolves around a “why” question: Why do some social contexts, some internal psychological states, and some individual differences modulate implicit attitudes and beliefs about in- and outgroups? Is a single psychological mechanism responsible for it or are multiple mechanisms responsible? Put differently, the primary purpose of this chapter is to use the distinction between automaticity and cognitive control to shed light on the conditions under which, and the mechanisms by which, implicit attitudes and beliefs about social groups change temporarily or chronically.

For purposes of this chapter, I refer to rapid judgments and evaluations made under time pressure as “implicit” because these responses are clearly driven, at least in part, by automatic processes when cognitive resources are limited. At the same time we now know that implicit attitude expressions are influenced, to varying extents, by the exertion of control. As the influence of controlled processes increases, judgments and evaluations become more intentional, effortful, and conscious.

The rest of this chapter is organized around two possible mechanisms likely to be responsible for the attenuation or exacerbation of implicit intergroup bias: changes in cognitive control and changes in the accessibility of automatic associations. I review existing research showing modulations in implicit intergroup judgments by linking each research finding to one of the two mechanisms. These links, of course, are speculative and are proposed here as predictions that need to be empirically tested in future research. Acquiring a better understanding of why implicit bias against outgroups is alleviated by some situations but not others, and by some goals, motivations, and emotions but not others, promises to provide traction in designing future interventions that might effectively tackle these subtle forms of bias.

INCREASING THE SALIENCE OF GROUP MEMBERSHIP INCREASES IMPLICIT BIAS BY ACTIVATING AUTOMATIC ASSOCIATIONS

A number of studies have found that increasing the salience of in- and outgroups magnifies implicit preference for ingroups and bias against outgroups. Some of these studies manipulated category salience by drawing perceivers' attention to target individuals' social identity (Macrae, Bodenhausen, Milne, & Calvini, 1999; Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997) or by drawing attention to perceivers' own social identity (Bohner, Seibler, Gonzalez, Haye, & Schmidt, 2007; Sassenberg & Wieber, 2005). Similarly, the presence of sufficient attentional resources has been shown to increase the activation of racial stereotypes after exposure to Asian or Black individuals compared to the absence of attentional resources (Gilbert & Hixon, 1991; Spencer, Fein, Wolfe, Fong, & Dunn, 1998).

Other studies drew attention to ingroup–outgroup distinctions by manipulating task goals or varying the exemplars used to represent particular social groups. For instance, when White participants

were instructed to attend to race (i.e., asked to classify White men and Black women along racial dimensions), they exhibited implicit preference for White men and bias against Black women. However, when participants were instructed to attend to gender (i.e., asked to classify the same individuals along the dimension of gender), they exhibited implicit preference for Black women and bias against White men (Mitchell, Nosek, & Banaji, 2003). In other words, attention to race or gender determined how individuals were categorized, which in turn influenced implicit evaluations of those individuals. The pervasive tendency to prefer White Americans only emerged when perceivers' attention was drawn to race, and a similar tendency to prefer women over men only emerged when perceivers' attention was drawn to gender. When an alternative social category membership was made salient, the same individuals were evaluated quite differently.

Category salience was also enhanced by varying individual representatives of a social group using stereotype-consistent versus stereotype-inconsistent members to represent the group. Implicit evaluations of outgroups tend to be significantly more negative if individual members fit the outgroup stereotype in terms of personality, social role, or physical appearance than if they are atypical. When individuals fit the stereotype or prototype of their group, more attention is drawn to category membership, which in turn evokes more implicit bias. For example, in the context of race, Mitchell et al. (2003) found that participants expressed strong implicit White preference when racial categories were represented with infamous Black individuals and famous White individuals. However, when the likeability of individual exemplars was reversed (famous Blacks and infamous Whites), implicit favoritism for Whites became nonsignificant (see Govan & Williams, 2004, for a similar effect). Similarly, individual outgroup members who fit the prototype of their group in terms of physical appearance tend to elicit more implicit negativity than others who do not fit the prototype. As a case in point, Black individuals with African facial features (darker complexion, fuller lips, broader nose) elicited more negative evaluations from White participants than Black individuals with less African facial features (light complexion, narrow lips and nose; Blair, Judd, & Fallman, 2004; Livingston & Brewer, 2002).

The effect of attention and category salience on implicit intergroup attitudes is not limited to known groups. A similar pattern of data emerges when fictitious groups are created in the laboratory. Drawing perceivers' attention to newly created groups produces implicit preference for individuals who are presented as ingroup members and bias against others who are presented as outgroup members (Ashburn-Nardo, Voils, & Monteith, 2001; Castelli, Zogmeister, Smith, & Arcuri, 2004; Otten & Wentura, 1999).

The common theme connecting all these studies is that they all drew participants' attention to particular types of category memberships (race, gender, etc.) which in turn probably activated default automatic evaluations associated with base categories (in the case of known groups) or created new associations (in the case of fictitious groups). Indirect evidence for this speculation comes from Payne et al. (2002), who found that drawing attention to the racial dimension of a speeded weapon identification task significantly increased race stereotypic errors compared to another condition where race was not emphasized prior to task performance. When participants' responses were disaggregated into automatic and controlled components (Jacoby, 1991), Payne and colleagues found that race-biased errors in identifying weapons were entirely driven by an increase in the automatic activation of racial stereotypes in the "race salient condition" compared to the "race not salient condition," whereas cognitive control did not change across the "race salient" versus "not salient" conditions. Taken together, this finding and the others already summarized suggest that increased attention to the social category membership of outgroup members and emphasis on ingroup-outgroup distinctions facilitate the activation of default automatic evaluations linked to ingroups versus outgroups: ingroup = good and outgroup = bad.

INCREASING THE SALIENCE OF COUNTERSTEREOTYPIC CUES DECREASES IMPLICIT BIAS BY ACTIVATING DIFFERENT POSITIVE ASSOCIATIONS

Social contexts that embody counterstereotypic or stereotypic cues have been known to significantly influence implicit evaluations and judgments of target group members seen in that context. For instance, exposure to African Americans in positive situations such as a family barbeque or church decreases implicit anti-Black bias relative to no-context controls, whereas exposure to the same individuals in negative situations such as a blighted inner-city street or in prison increases implicit anti-Black bias (Barden, Maddux, Petty, & Brewer, 2004; Rudman & Lee, 2002; Wittenbrink, Judd, & Park, 2001). Similarly, situations that make salient the positive cultural and historical contributions of Arab societies decrease implicit anti-Arab bias relative to a neutral context, whereas situations that make salient news about terrorism increase anti-Arab bias relative to a neutral context (Park, Felix, & Lee, 2007).

These effects are not limited to background features of social situations. Other cues in the foreground of social situations also modulate implicit evaluations and judgments. Situations that primed exposure to counterstereotypic members of disadvantaged groups prior to the measurement of implicit attitudes and beliefs revealed a substantial decline in implicit negativity against outgroups (e.g., elderly, African Americans, gay men; Dasgupta & Greenwald, 2001; Dasgupta & Rivera, 2008) and implicit stereotyping of ingroups (e.g., women; Dasgupta & Asgari, 2004). Such reduction of implicit bias was particularly evident for individuals whose everyday social environments provided little opportunity for close contact with outgroups. Others who had a great deal of prior contact showed less outgroup bias regardless of the situational manipulation in the laboratory (Dasgupta & Rivera, 2008). The influence of counterstereotypic individuals on the reduction of implicit bias has been shown to occur and endure even in year-long longitudinal studies (Asgari, Dasgupta, & Gilbert-Cote, 2008; Dasgupta & Asgari, 2004) and is not limited to a brief moment in the laboratory. There is, however, some sobering evidence suggesting that increased contact with outgroup members appears to elicit less of an impact on the attitudes of individuals who belong to high-status groups (White Americans in the United States and Christians in Lebanon) compared to those who belong to lower status groups (Black Americans in the United States and Muslims in Lebanon, respectively; Henry & Hardin, 2006).

Another situational cue that modulates implicit attitudes is perceivers' social role relative to their interaction partners. Specifically, White participants or male participants who anticipate an impending cross-race or cross-gender interaction where their interaction partner is in a superior (counterstereotypic) role exhibit less outgroup bias than others who anticipate interacting with an outgroup member who is in a subordinate (stereotypic) role (Richeson & Ambady, 2001, 2003). At the same time, however, cross-gender interactions with a woman in a superior (rather than subordinate) role cause men to implicitly compensate and stereotype themselves as more masculine (McCall & Dasgupta, 2007).

Such situation-driven changes in implicit bias are likely to be elicited by the activation of different mental associations linking social groups to counterstereotypic attributes (see Gawronski & Bodenhausen, 2006). Although stereotypic associations have greater accessibility in default situations or decontextualized experimental situations, the introduction of counterstereotypic cues (background features of situations, social roles of interaction partners, counterstereotypic individuals) enhances the accessibility of other (counterstereotypic) associations linked to target groups. Such cues may also suppress the accessibility of stereotypic associations if stereotypic and counterstereotypic attributes are perceived to be bipolar constructs that cannot be activated simultaneously (see Greenwald et al., 2002; Heider, 1958). Moreover, long-term immersion in counterstereotypic social contexts may reduce the default accessibility of stereotypes or enhance the chronic accessibility of counterstereotypes, thereby decreasing the likelihood of biased automatic judgments and evaluations in the future.

SPECIFIC MOTIVATIONS CAN INCREASE OR DECREASE IMPLICIT BIAS BY CHANGING IN COGNITIVE CONTROL

Although rapid evaluations and judgments under extreme time pressure have been typically assumed to bypass motivational processes, accumulating evidence has begun to reveal that chronic individual differences in motivation as well as situationally triggered motivation modulate implicit judgments of in- and outgroups. The specific source of motivation may be perceivers' emotional state, desire to protect self-esteem or group esteem, motivation to control prejudice, or a generalized capacity for executive control. Moreover, depending on the specific circumstance, motivation may decrease or increase implicit bias.

EMOTION AS A SOURCE OF MOTIVATION

In one program of research we have found that when people experience certain discrete emotions (e.g., anger, disgust) that are associated with motivations to aggress or avoid, the residue of the emotion spills over from the original source to bias implicit evaluations of real and fictitious groups, even when the emotion-inducing source is unrelated to social groups (Dasgupta, DeSteno, Pressman, Williams, & Hunsinger, Yogeeswaran, & Ashby, 2007; DeSteno, Dasgupta, Bartlett, & Caidric, 2004). Interestingly, the biasing effect of emotion on outgroup evaluations only occurs for intergroup negative emotions (e.g., anger and disgust), not all negative emotions (e.g. sadness).

Moreover, although both anger and disgust are capable of creating implicit bias against previously neutral and unknown outgroups, these two emotions have differential effects on appraisals of known outgroups. Specifically, incidental feelings of anger (but not disgust) exacerbate implicit bias against Arabs and incidental feelings of disgust (but not anger) increase implicit bias against gays and lesbians (Dasgupta et al., 2007). We propose that because disgust is elicited by physical or moral contaminants, and because gays and lesbians are perceived to violate mainstream moral values about "appropriate" sexual behavior (Cottrell & Neuberg, 2005; Herek, 1996; Mosher & O'Grady, 1979; Nussbaum, 1999), incidental feelings of disgust are experienced as applicable to this group and thus have a spillover effect.

Similarly, because anger is elicited when people confront obstacles, and experience threats to their economic resources, freedoms, and rights (Cottrell & Neuberg, 2005; Fiske, Cuddy, Glick, & Xu, 2002; Mackie, Devos, & Smith, 2000; Smith, 1993), and because contemporary stereotypes of Arabs include anger-relevant attributes (Park et al., 2007), when people feel angry for incidental reasons, that emotion spills over into appraisals of Arabs. Our data showed that anger increased anti-Arab bias by depleting cognitive control (Dasgupta et al., 2007). Specifically, angry participants' erroneous evaluations were driven by a significant reduction in controlled processing compared to others who felt neutral or disgusted.

SELF-IMAGE THREAT AND SOCIAL IDENTITY THREAT AS A SOURCE OF MOTIVATION

Motivation to maintain a positive self-image or ingroup image also influences implicit attitudes toward outgroups. Self-threat and social identity threat have been found to increase implicit stereotyping and prejudice (Gonsalkorale, Carlisle, & von Hippel, 2007; Spencer et al., 1998) and to enhance collective self-esteem postjudgment (Gonsalkorale et al., 2007). The implication here is that threats to self-esteem and group esteem motivate people to derogate a target outgroup as a way of recovering positive self-regard or ingroup regard.

Similarly, criticism from an outgroup member increases implicit outgroup bias, whereas praise decreases outgroup bias. Specifically, White participants who had received praise from a Black or Asian person in a higher status role (e.g., manager, doctor, or experimenter) subsequently exhibited less implicit stereotypes about the relevant outgroup compared to others who had received criticism from the same person (Sinclair & Kunda, 1999). These data imply that praise validates perceivers'

self-esteem and motivates them to like the praise-bearing messenger (outgroup member) and others in his or her group, whereas criticism invalidates their positive self-esteem and motivates them to dislike the criticism-bearing messenger and others in that outgroup.

In the studies mentioned earlier, self-esteem threat and social identity threat may have operated by modulating cognitive control. Specifically, threat may have decreased the motivation to be accurate by exerting control, which in turn increased stereotypic responses, whereas praise may have increased the motivation to be accurate by exerting control, which reduced stereotypic responses. In addition, self-threat and social identity threat may have also increased the accessibility of negative associations linked to the particular outgroup.

PROMOTION AND PREVENTION FOCUS AS A SOURCE OF MOTIVATION

Recent research has found that individuals' regulatory states (i.e., whether they are oriented toward accruing gains [promotion focus] or avoiding losses [prevention focus] influence their implicit attitudes toward groups that vary in power (Sassenberg, Jonas, Shah, & Brazy, 2007). Specifically, higher power ingroups are more likely to meet the regulatory needs of individuals with a promotion focus than a prevention focus because such ingroups give promotion-focused individuals the opportunity to pursue their ideals (e.g., to seek achievement, nurturance, etc.). Conversely, lower power ingroups meet the regulatory needs of individuals with a prevention focus more than a promotion focus because such ingroups allow their members to pursue behaviors that focus on safety and security to prevent losses. As predicted, Sassenberg and colleagues found that high-power ingroups were more implicitly favored by their members who had a promotion focus rather than a prevention focus, whereas lower power ingroups were more implicitly favored by their members who had a prevention focus rather than promotion focus.

Regulatory focus may have influenced cognitive control. Specifically, promotion focus may have increased individuals' motivation to attend to the desirable qualities of high-power ingroups, whereas prevention focus may have increased their motivation to attend to the positive qualities of lower power ingroups. In addition, changes in attention consistent with regulatory focus may have also changed the accessibility of particular attributes linked to in- and outgroups.

SOCIAL NORMS AS A SOURCE OF MOTIVATION

A classic finding in social psychology is that people tend to conform to norms in their social environment because they are motivated to be liked by, and be similar to, their peer group (Asch, 1955, 1956). As such, individuals who become aware that their opinions are out of sync with their peers tend to shift their attitudes toward the peer group. This normative influence is also apparent when it comes to implicit attitudes. Sechrist and Stangor (2001) found that participants' implicit beliefs about African Americans became less stereotypic if they discovered that their peer group was more egalitarian than themselves compared to a situation in which they had no information about peer opinion. However, participants' beliefs became more stereotypic if they discovered that their peer group was less egalitarian than themselves compared to "no information" controls. As in the case of Asch's famous conformity experiments, these shifts in implicit attitudes point to the role of normative influence; they suggest that awareness of social norms increased participants' motivation to be similar to their peers, which subsequently increased the degree to which they attempted to control and modify their outgroup evaluations to fit in.

MOTIVATION TO CONTROL PREJUDICE

Aside from situationally triggered motivations of the sort already described, individual differences in chronic motivation also affect implicit attitudes. For example, individual differences in motivation to control prejudice are known to moderate implicit racial attitudes in systematic ways.

Correlational research has found that people who are highly motivated to control prejudice show less implicit race bias than their less motivated peers. Moreover, although lower race bias is correlated with greater cognitive control, it is not correlated with race-biased automatic associations (Payne, 2005). Experimentally induced motivation (i.e., being reminded of one's past race-based transgressions) also decreases bias in subsequent behavior especially among people who are implicitly prejudiced yet explicitly egalitarian (i.e., aversive racists; Son Hing, Li, & Zanna, 2002; also see Son Hing, Chung-Yan, Grunfeld, Robichaud, & Zanna, 2005).

In some cases, however, high motivation to control prejudice can backfire and increase race bias if people are explicitly made aware that the task they are about to complete reveals racial prejudice for the majority of test-takers (Frantz, Cuddy, Burnett, Ray, & Hart, 2004; but see Nier, 2005). As Frantz and colleagues noted, this ironic effect may occur because individuals with high motivation to control prejudice are likely to be concerned about their own unintentional bias when made aware of the nature of the task; this concern in turn may interfere with their usual capacity to respond accurately (see Richeson, Baird, & Gordon, 2003; Richeson & Shelton, 2003). In addition to reduced cognitive control, concern about appearing prejudiced may also lead people to monitor stereotypic thoughts, which may inadvertently increase the accessibility of racial stereotypes revealed in subsequent judgments (see Wegner, 1994).

GLOBAL EXECUTIVE CONTROL AS A SOURCE OF ACCURACY MOTIVATION

Behavioral expressions of implicit stereotypes and prejudice may also be shaped by individual differences in people's capacity for executive control in general—this refers to basic attentional capacities that allow people to selectively attend to information that is relevant to the task goals at hand and simultaneously screen out other information that is irrelevant to those goals (Macrae, Bodenhausen, Schloerscheidt, & Milne, 1999; Payne, 2005). With one exception (Payne, 2005) there is little research examining the relationship between global executive control and implicit stereotypes and prejudice. Payne (2005) reported correlational evidence showing that individuals who exhibited better executive control on a task unrelated to social groups also showed more controlled processing and less bias against African Americans on a number of race-based speeded tasks (e.g., weapons identification task, evaluative priming task, and Implicit Association Test). Moreover, global executive control was uncorrelated with motivation to control prejudice.

Indirect evidence for the benefit of global executive control comes from research that examines the role of practice and training (i.e., improvement in executive control) on the accuracy of intergroup judgments. Specifically, Correll et al. (2007) compared police officers' and community members' decisions to shoot (or not shoot) Black and White men in a law enforcement simulation. They found that trained police officers outperformed community members in terms of overall speed and accuracy. Moreover, whereas community respondents used a relaxed decision criterion to shoot Black compared to White targets (thereby making more race-biased errors), police officers used a stricter criterion. However, both samples exhibited robust race bias in response speed. These data suggest that training (and presumably increased executive control as a result of it) encourages the selection of a stricter decision criterion before shooting, although it does not affect the speed with which stereotype-incongruent targets are processed.

Given the suggested benefit of global executive control, one avenue for future research is an investigation of whether such control moderates the effectiveness of social environments; that is, when immersed in counterstereotypic social environments, do individuals with better executive control exhibit a steeper decline in implicit bias compared to their peers who have less executive control?

THE INFLUENCE OF CONTEXTUAL CUES IS MODERATED BY INDIVIDUAL AND GROUP DIFFERENCES: THE COMBINED RULE OF AUTOMATIC ASSOCIATIONS AND COGNITIVE CONTROL

Initial research on the flexibility of implicit attitudes focused either on the role of social contexts or on the role of motivational processes, but not both. Collectively, this research illustrated that different types of situational cues and internal states change implicit appraisals of in- and outgroups. But what about the combined effect of social context and individual difference? Only a handful of studies have examined this question; they have found that the effect of social contexts on implicit attitudes is dependent on: (a) people's motivation to control prejudice (Maddux, Barden, Brewer, & Petty, 2005); (b) individual differences in social dominance (Pratto & Shih, 2000); (c) chronic beliefs in a dangerous world (Schaller, Park, & Mueller, 2003); and (d) perceivers' ingroup membership (Kühnen, Schießl, and Bauer, 2001; Lowery, Hardin, & Sinclair, 2001).

For example, when African Americans and White Americans were seen in negative social contexts (e.g., jail, foggy road), White perceivers' reactions depended on individual differences in their motivation to control prejudice. Whereas individuals who were not motivated to control prejudice exhibited implicit ingroup preference for Whites, others who were highly motivated to control prejudice showed the reverse effect: they exhibited outgroup preference for Blacks (Maddux et al., 2005). These race-based evaluations were driven by slower reaction to stereotypic (Black-negative) associations. Analogously, Schaller and colleagues (2003) found that participants who harbored chronic beliefs that the world is a dangerous place showed more implicit bias against African Americans when placed in a potentially ominous situation (darkened room) compared to a less ominous situation (well-lit room), but others who did not share these chronic beliefs showed no changes in implicit attitudes as a function of context. Similarly, individual differences in social dominance modulated the magnitude of outgroup bias when perceivers' ingroup status was threatened. Individuals high in social dominance exhibited more ingroup favoritism and outgroup derogation than others who were low in social dominance, but only when the high status of their ingroup was called into question (Pratto & Shih, 2000).

Perceivers' group membership also moderates the effect of counterstereotypic social cues on implicit racial attitudes. For example, Lowery et al. (2001) found that whereas White participants exhibited substantially less implicit preference for Whites after interacting with a Black compared to a White experimenter, Asian participants' implicit racial attitudes did not vary as a function of experimenter race. Along the same lines, Kühnen et al. (2001) compared the impact of increasing the salience of ingroup identity on implicit intergroup attitudes. They predicted and found that when people's ingroup was positively stereotyped (e.g., being West German), increasing the salience of group identity exacerbated implicit ingroup favoritism but when people's ingroup was negatively stereotyped (e.g., being East German), increasing the salience of group identity attenuated ingroup favoritism.

The common theme running through these findings is that some individual- and group-level variables influence the degree of motivation and control people are willing or able to invest in ensuring that their rapid responses to in- and outgroups are accurate. Those who are highly motivated to control prejudice, those who do not believe that the world is a dangerous place, and those who belong to an advantaged group that historically has been the agent of discrimination, may all be particularly invested in monitoring and controlling their evaluative reactions in a racial context to override potentially biased automatic reactions. Thus, although stereotypic contextual cues enhance the accessibility of automatic stereotypes for everybody, increased motivation to control prejudice may trigger efforts to exert control, thereby overriding potential biases even in time-pressured situations.

LEARNING AND UNLEARNING IMPLICIT ATTITUDES: THE ROLE OF COGNITIVE CONTROL AND AUTOMATIC ASSOCIATIONS

Attitude change in general, and prejudice reduction in particular, requires unlearning old attitudes and learning new ones. Prejudice change is typically thought to occur when people consciously invest effort to reconsider old attitudes in light of new information or when they learn to suppress thoughts that are seen as invalid or inappropriate and replace them with new thoughts. To what extent do such conscious unlearning and learning strategies influence implicit attitudes toward social groups? If such strategies have an impact, do they change the accessibility of preexisting associations, change cognitive control, or both? Does the type of target group influence the effectiveness of learning and unlearning? Does expertise influence learning and unlearning? Recent research has begun to address some of these questions.

LEARNING AND UNLEARNING ATTITUDES BY MERE INSTRUCTION VERSUS CONCRETE STRATEGIES

Research shows that even when people are made aware that their attitudes toward disadvantaged groups are being measured, and they are explicitly instructed to respond in an egalitarian manner, their implicit attitudes continue to exhibit bias against African Americans and gay men—although their explicit attitudes become less biased (Banse, Seise, & Zerbes, 2001; Kim, 2003). In fact, explicit instructions to suppress preexisting stereotypes have been known to produce an ironic effect in some studies by exacerbating the activation of implicit stereotypes (Galinsky & Moskowitz, 2000; Macrae et al., 1994). This is probably because bias suppression instructions actively draw people's attention to outgroup membership and enhance the accessibility of default stereotypic associations.

However, when instructions offer a concrete strategy that people can use to consciously override implicit bias, outgroup bias is attenuated. For example, participants who received a specific implementation intention to avoid gender stereotypes from biasing their judgments of an individual were able to control and reduce implicit stereotyping (Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). Similarly, participants who received concrete instructions on how to modify their rapid race-based responses on a reaction time task were able to reduce implicit race bias (Kim, 2003). Likewise, others who were told to expect counterstereotypic information when they thought of women and men also showed reduced implicit gender stereotyping (Blair & Banaji, 1996).

Another type of concrete strategy that produces beneficial effects involves instructing people to mentally elaborate on the positive or counterstereotypic qualities of outgroups. Cognitive elaboration is likely to enhance the accessibility of counterstereotypic attributes associated with target groups that emerge with greater strength in subsequent implicit evaluations. For example, a number of studies have found that when people were instructed to imagine and describe women with counterstereotypic qualities they subsequently showed reduced implicit gender stereotyping compared with others who were not asked to engage in such cognitive elaboration or, alternatively, who were asked to describe women with stereotypic qualities (Blair, Ma, & Lenton, 2001; Carpenter, 2001). Cognitive elaboration also works somewhat indirectly by making salient the subjective ease of recalling well-liked members of an outgroup or the difficulty of recalling disliked members of the same group (Gawronski & Bodenhausen, 2005).

Just as instructions to think about the positive qualities of disfavored outgroups significantly attenuate previously ingrained implicit bias, the subsequent generation of negative thoughts increases outgroup bias again (Akalis & Banaji, 2008). Interestingly, the more skilled people are at mental discipline through yoga and meditation, the better they are at reducing implicit bias even when given a fairly open-ended instruction to reduce their prejudice “by whatever mental means possible” or to generate feelings of compassion and kindness toward a particularly disliked outgroup (Akalis & Banaji, 2008). Activating a creative mindset has a similar effect by reducing implicit stereotyping (Sassenberg & Moskowitz, 2005). In both these cases, experimental instructions may spontaneously increase elaboration about the positive qualities possessed by outgroup members;

alternatively, they may increase positive affect directed at the outgroup without elaboration about the specific reasons.

Other types of mental elaboration have been shown to attenuate implicit outgroup bias: People who read about multiculturalism and elaborated on the benefits of celebrating interethnic diversity subsequently exhibited less implicit race bias than others who read about color-blind values and elaborated on the benefits of ignoring group membership (Richeson & Nussbaum, 2004; but see Smyth & Nosek, 2007). Similarly, when college students thought about and elaborated on the benefits of diversity during a semester-long class on intergroup relations and diversity, they subsequently showed reduced implicit race bias in a pretest–posttest field study (Rudman, Ashmore, & Gary, 2001).

However, cognitive elaboration can sometimes backfire on perceivers who are ambivalently prejudiced. Specifically, Maio, Haddock, Watt, and Hewstone (in press) found that racially ambivalent participants (but not nonambivalent participants) who encountered antiracism advertisements exhibited an increase in implicit race bias when the ad presented weak arguments, suggesting that these participants were motivated to scrutinize the quality of the advertisement carefully and found it wanting.

Taken together, a common theme underlying the studies just mentioned is that when people engage in cognitive elaboration exercises that increase the salience of counterstereotypes or that encourage a different way of thinking, such directed thinking increases the accessibility of counterstereotypic associations linked to outgroups, which in turn temporarily alleviates implicit bias against outgroups. In other words, even when implicit judgments are made in highly time-pressured situations they can be debiased if people acquire concrete strategies that allow them to override and modify their automatic responses. These strategies function as detailed action plans on how to exert control whereas the mere instruction to avoid bias is clearly not sufficient and sometimes even counterproductive. The only time cognitive elaboration boomerangs and increases implicit prejudice is when perceivers are ambivalently biased to begin with and they encounter information that presents weak arguments in favor of egalitarianism.

So far, there appear to be two exceptions to the failure of “mere instruction” strategies: Mere instructions to decrease implicit bias appear to work for: (a) a group of participants with special training in mental discipline; and (b) fictitious attitude objects (social groups) created in the laboratory rather than real ones. In the first case, people who are skilled at mental discipline through the practice of yoga and meditation appear to be particularly good at decreasing implicit bias even when they are given a fairly open-ended instruction to reduce their prejudice “by whatever mental means possible” or when they are instructed to generate feelings of compassion and kindness toward a particularly disliked outgroup (Akalis & Banaji, 2008).

In the second case, the simple instruction to imagine a fictitious group as possessing positive (or negative) qualities appears to be sufficient to create new implicit attitudes that are consistent with the imagined quality (Gregg, Seibt, & Banaji, 2006). Although a mere suggestion can create implicit attitudes toward unknown groups from thin air, once these attitudes are formed they are more difficult to unlearn compared to their explicit counterparts. For example, Gregg and colleagues (2006) found that although implicit and explicit attitudes toward hypothetical social groups were influenced by the valence of information initially presented about these groups, when participants were later told that the initial information was false and the truth was actually the opposite of what they had initially learned, this new information either had no unlearning effect (Experiment 3) or a weak effect (Experiment 4) on implicit attitudes; however, it had a strong unlearning effect on explicit attitudes.

Conceptually similar results have been obtained for implicit attitudes toward unknown individuals (Petty et al., 2006; Rydell & McConnell, in press). For example, Petty and colleagues (2006) found that once initial attitudes had been formed about unfamiliar individuals, later invalidation of the initial information reversed explicit attitudes toward those individuals but could not reverse (albeit neutralized) implicit attitudes. The take-home message of these studies is that once people

have acquired knowledge and opinions about groups and individuals, simply informing them that their knowledge is false does not allow them to reverse their responses through sheer willpower and control, nor does it change the accessibility of underlying mental associations (see Gawronski & Bodenhausen, 2006).

LEARNING AND UNLEARNING ATTITUDES BY EXTENDED TRAINING

An alternative way of modifying preexisting attitudes about known groups is to allow opportunities for extended training (rather than providing simple instruction) to enhance the accessibility of counterstereotypic associations about historically stereotyped groups. Several studies have found that when stimuli representing one social group (e.g., old people, African Americans, Asian, a fictitious group) were repeatedly paired with positive attributes, and stimuli representing a contrasting group were repeatedly paired with negative attributes (e.g., young people, White Americans, European), such extended training changed implicit intergroup evaluations (Gawronski, LeBel, Heilpern, & Wilbur, 2007, Experiment 2; Glaser, 1999, Experiment 2; Karpinski & Hilton, 2001; Olson & Fazio, 2006). Similarly, when participants were extensively trained to negate stereotypes (Kawakami et al., 2000) or affirm counterstereotypes (Gawronski et al., 2008) their implicit evaluations of target groups became less stereotypic compared to others who were trained to affirm stereotypes or who received no training at all. Recent research suggests that affirming counterstereotypes is a more effective way of reducing implicit stereotypes than negating stereotypes (Gawronski et al., 2007). Clearly, all these interventions work by temporarily changing the accessibility of underlying automatic associations about specific groups. However, they also require that perceivers be willing to engage in fairly effortful and intentional training processes and are thus quite different from situational interventions where incidental cues in the social context provide an indirect means of changing automatic associations about social groups.

CONCLUSION AND NEW DIRECTIONS

Although implicit prejudices and preferences are pervasive and reflective of stable societal inequalities, at an individual level these attitudes are remarkably flexible. Even when individuals' cognitive resources are depleted, their response time is tightly constrained, or they have limited awareness of the implications of their actions, individuals' behaviors continue to be remarkably pliable. This chapter highlighted two of the possible mechanisms that are likely to be responsible for driving the attenuation versus exacerbation of implicit bias depending on the situation: changes in automatic associations about social groups and changes in cognitive control over one's behavior. Generally speaking, situations that draw attention to social group membership, stereotypic or counterstereotypic group members, and status differences in social roles, and situations that offer extended training with counterstereotypic individuals are likely to change the magnitude of implicit bias by influencing the accessibility of particular group-related qualities. In comparison, situations that evoke specific motivations due to perceivers' emotions, normative influences, self-image or social identity threat, and particular individual differences are likely to modulate the magnitude of implicit bias via different route—by guiding the ebb and flow of cognitive control. Because little empirical research has directly pinpointed the underlying mechanisms driving specific shifts in implicit attitudes toward in- and outgroups, the goal of this chapter was to generate process-oriented hypotheses for future research based on indirect evidence. A deeper knowledge of the processes by which implicit intergroup attitudes change promises to inform other important questions such as how chronic versus temporary these changes are. And when might particular bias reduction strategies be translated from laboratory paradigms to real-world interventions?

REFERENCES

- Akalis, S., & Banaji, M. R. (2008). *Do-it-yourself mental makeovers: Self-generated associations shift implicit references*. Manuscript in preparation, Harvard University, Cambridge, MA.
- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*, 31–35.
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs*, *70*, 70.
- Asgari, S., Dasgupta, N., & Gilbert-Cote, N. (2008). *The influence of female leaders on women's implicit self-concept: Investigations in the lab and field*. Manuscript under review.
- Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology*, *81*, 789–799.
- Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, III & J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.
- Banaji, M. R., & Greenwald, A. G. (1994). Implicit stereotyping and prejudice. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium* (Vol. 7, pp. 55–76). Mahwah, NJ: Erlbaum.
- Banse, R., Seise, J., & Zerbos, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, *48*, 145–160.
- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5–22.
- Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Erlbaum.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life: Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.
- 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.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–261.
- Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*, 1142–1163.
- Blair, I. V., Judd, C. A., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, *87*, 763–778.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841.
- Bohner, G., Seibler, F., Gonzalez, R., Haye, A., & Schmidt, E. A. (2007). *Changing implicit group evaluations by priming social identities: The case of people with dual (Turkish/German) national identity*. Manuscript under review.
- Carpenter, S. J. (2001). *Implicit gender attitudes*. Unpublished manuscript, Yale University, New Haven, CT.
- Castelli, L., Zogmeister, C., Smith, E. R., & Arcuri, L. (2004). On the automatic evaluation of social exemplars. *Journal of Personality and Social Psychology*, *86*, 373–387.
- Conroy, 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.
- 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 and Social Psychology*, *92*, 1006–1023.
- Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice.” *Journal of Personality and Social Psychology*, *88*, 770–789.
- Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, *17*, 143–169.
- Dasgupta, N. (in press). Color lines in the mind: Unconscious prejudice, discriminatory behavior, and the potential for change. In A. Grant-Thomas & G. Orfield (Eds.), *21st century color lines: Exploring the frontiers of America's multicultural future*.
- Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*, 642–658.

- Dasgupta, N., DeSteno, D. A., Pressman, A., Williams, L., Yogeeswaran, K., Hunsinger, M., & Ashby, J. (2007). A tale of emotion specificity: *Behavioral and ERP evidence for the effect of incidental emotions in implicit prejudice*. Manuscript in preparation.
- 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.
- Dasgupta, N., & Rivera, L. M. (2008). When social context matters: The influence of long-term contact and short-term exposure to admired outgroup members on implicit attitudes and behavioral intentions. *Social Cognition, 26*, 54–66.
- DeSteno, D. A., Dasgupta, N., Bartlett, M. Y., & Caidric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes. *Psychological Science, 15*, 319–324.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.
- Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60*, 817–830.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878–902.
- Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race Implicit Association Test as a stereotype threat experience. *Personality and Social Psychology Bulletin, 30*, 1611–1624.
- Gaertner, S. L., & Dovidio, J. F. (1986). An aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). New York: Academic.
- Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology, 78*, 708–724.
- Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology, 89*, 672–685.
- 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., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When “just say no” is not enough: Affirmation vs. negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology, 44*(2), 370–377.
- Gawronski, B., LeBel, E. P., Heilpern, K., & Wilbur, C. J. (2007). *Understanding the relative robustness of implicit and explicit attitudes: A test of the associative-propositional evaluation model*. Manuscript under review.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.
- Glaser, J. C. (1999). *The relation between stereotyping and prejudice: Measures of newly formed automatic associations*. Unpublished manuscript, Yale University, New Haven, CT.
- Gonsalkorale, K., Carlisle, K., & von Hippel, W. (2007). Intergroup threat increases implicit stereotyping. *International Journal of Psychology and Psychological Therapy, 1*, 189–200.
- Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology, 40*, 357–365.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3–25.
- Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*, 1–20.
- Heider, F. (1958). *The psychology of interpersonal relations*. Hoboken, NJ: Wiley.
- Henry, P. J., & Hardin, C. D. (2006). The contact hypothesis revisited: Status bias in the reduction of implicit prejudice in the United States and Lebanon. *Psychological Science, 17*, 862–868.
- Herek, G. M. (1996). Heterosexism and homophobia. In R. P. Cabaj & T. S. Stein (Eds.), *Textbook of homosexuality and mental health* (pp. 101–113). Washington, DC: American Psychiatric Association.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774–788.
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology, 55*, 893–905.
- Katz, I., Wackenhut, J., & Hass, R. G. (1986). Racial ambivalence, value duality, and behavior. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 35–60). New York: Academic.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effect of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78*, 871–888.
- Kihlstrom, J. F. (1990). The psychological unconscious. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 445–464). New York: Guilford.
- Kim, D.-Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly, 66*, 83–96.
- Kühnen, U., Schießl, M., & Bauer, N. (2001). How robust is the IAT? Measuring and manipulating implicit attitudes of East- and West-Germans. *Zeitschrift für Experimentelle Psychologie, 48*, 135–144.
- Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5–18.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*, 842–855.
- Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology, 79*, 602–616.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Calvini, G. (1999). Seeing more than we can know: Visual attention and category activation. *Journal of Experimental Social Psychology, 35*, 590–602.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67*, 808–817.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology, 33*, 471–489.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Wheeler, V. (1996). On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory. *Social Cognition, 14*, 1–20.
- Macrae, C. N., Bodenhausen, G. V., Schloerscheidt, A., & Milne, A. B. (1999). Tales of the unexpected: Executive function and person perception. *Journal of Personality and Social Psychology, 76*, 200–213.
- Maddux, W. W., Barden, J., Brewer, M. B., & Petty, R. E. (2005). Saying no to negativity: The effects of context and motivation to control prejudice on automatic evaluative responses. *Journal of Experimental Social Psychology, 41*, 19–35.
- Maio, G. R., Haddock, G., Watt, S. E., & Hewstone, M. (in press). Implicit measures and applied contexts: An illustrative examination of anti-racism advertising. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new wave of implicit measures*.
- McCall, C., & Dasgupta, N. (2007). The malleability of men's gender self-concepts. *Self and Identity, 6*(2–3), 173–188.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*, 455–469.
- Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65*, 469–485.
- Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed vs. other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology, 64*, 198–210.
- Monteith, M. J., Zuwerink, J. R., & Devine, P. G. (1994). Prejudice and prejudice reduction: Classic challenges, contemporary approaches. In P. G. Devine, D. L. Hamilton, & T. M. Ostrom (Eds.), *Social cognition: Impact on social psychology* (pp. 323–345). San Diego, CA: Academic.
- Mosher, D. L., & O'Grady, K. E. (1979). Homosexual threat, negative attitudes toward masturbation, sex guilt, and males' sexual and affective reactions to explicit sexual films. *Journal of Consulting and Clinical Psychology, 47*, 860–873.
- Moskowitz, G. B., Gollwitzer, P., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*, 167–184.
- Myrdal, G. (1944). *An American dilemma; The Negro problem and modern democracy*. New York: Harper.

- Nier, J. A. (2005). How dissociated are implicit and explicit racial attitudes? A bogus pipeline approach. *Group Processes and Intergroup Relations*, 8, 39–52.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Nussbaum, M. C. (1999). “Secret sewers of vice”: Disgust, bodies, and the law. In S. A. Bandes (Ed.), *The passions of law* (pp. 19–62). New York: New York University Press.
- Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, 32, 421–433.
- Otten, S., & Wentura, D. (1999). About the impact of automaticity in the minimal group paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology*, 29, 1049–1071.
- Park, J., Felix, K., & Lee, G. (2007). Implicit attitudes toward Arab-Muslims and the moderating effects of social information. *Basic and Applied Social Psychology*, 29, 35–45.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181–192.
- Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, 89, 488–503.
- 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.
- Pettigrew, T., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90, 751–783.
- Petty, R. E., Tormala, Z. L., Brinol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology*, 90, 21–41.
- Pratto, F., & Shih, F. (2000). Social dominance orientation and group context in implicit group prejudice. *Psychological Science*, 11(6), 515–518.
- Richeson, J. A., & Ambady, N. (2001). Who’s in charge? Effects of situational roles on automatic gender bias. *Sex Roles*, 44, 493–512.
- Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology*, 39, 177–183.
- Richeson, J. A., Baird, A. A., & Gordon, H. L. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, 6, 1323–1328.
- Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, 40, 417–423.
- Richeson, J. L., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, 14, 287–290.
- Rockeach, M. (1973). *The nature of human values*. New York: The Free Press.
- Rudman, L. A., Ashmore, R. D., & Gary, M. (2001). “Unlearning” automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, 81, 856–868.
- Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes and Intergroup Relations*, 5, 133–150.
- Rydell, R. J., & McConnell, A. R. (in press). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*.
- Sassenberg, K., Jonas, K. A., Shah, J. Y., & Brazy, P. C. (2007). Why some groups just feel better: The regulatory fit of group power. *Journal of Personality and Social Psychology*, 92, 249–267.
- Sassenberg, K., & Moskowitz, G. B. (2005). Don’t stereotype, just think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology*, 41, 506–514.
- Sassenberg, K., & Wieber, F. (2005). Don’t ignore the other half: The impact of ingroup identification on implicit measures of prejudice. *European Journal of Social Psychology*, 35, 621–632.
- Schaller, M., Park, J. H., & Mueller, A. (2003). Fear of the dark: Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality and Social Psychology Bulletin*, 29, 637–649.
- Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology*, 80, 645–654.
- Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, 77, 885–904.
- Smith, E. R. (1993). Social identity and social emotions: Toward new conceptualizations of prejudice. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 297–315). San Diego, CA: Academic.

- 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.
- Smyth, F., & Nosek, B. A. (2007, January). *The impact of multiculturalism versus color-blindness on racial bias revisited*. Paper presented at the annual meeting of the Society for Personality and Social Psychology, Memphis, TN.
- Son Hing, L. S., Chung-Yan, G. A., Grunfeld, R., Robichaud, L. K., & Zanna, M. P. (2005). Exploring the discrepancy between implicit and explicit prejudice: A test of aversive racism theory. In J. P. Forgas, K. Williams, & S. Latham (Eds.), *Social motivation: Conscious and unconscious processes. Vol. 5. The Sydney Symposium of Social Psychology* (pp. 274–293). New York: Psychology Press.
- Son Hing, L. S., Li, W., & Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial response among aversive racists. *Journal of Experimental Social Psychology*, *38*, 71–77.
- Spencer, S. J., Fein, S., Wolfe, C. T., Fong, G., & Dunn, M. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*, 1139–1152.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.
- Tropp, L. R., & Bianchi, R. A. (2006). Valuing diversity and interest in intergroup contact. *Journal of Social Issues*, *62*, 533–551.
- Tropp, L. R., Stout, A. M., Boatswain, C., Wright, S., & Pettigrew, T. F. (2006). Trust and acceptance in response to references to group membership: Minority and majority perspectives on cross-group interactions. *Journal of Applied Social Psychology*, *36*, 769–794.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34–52.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.