Attitudes

Insights from the New Implicit Measures

Edited by

Richard E. Petty • Russell H. Fazio • Pablo Briñol

Psychology Press
Taylor & Francis Group
New York Hove

2

Implicit and Explicit Measures of Attitudes

The Perspective of the MODE Model

Michael A. Olson
Russell H. Fazio

Introduction

Strong or weak, extreme or mild, certain or uncertain, attitudes are as richly diverse as the judgments and behavior they often influence. Years of systematic study of these attitudinal qualities have tackled the age-old questions of when and how attitudes relate to behavior. What sorts of attitudes relate reliably to judgments and behavior? Under what conditions is attitude–behavior correspondence most likely to occur? By what processes do attitudes exert their influence? The MODE model (Motivation and Opportunity as Determinants of the attitude-behavior relation) was developed to address these historical and fundamental questions (Fazio, 1990; Fazio & Towles-Schwen, 1999). We will describe the tenets of the model in this chapter, and in so doing illuminate the multiple paths from attitude to behavior delineated by the MODE model.

However, in addition to these historical issues, the present volume also tackles a more contemporary set of questions. Premised on the potentially important distinction between implicit and explicit attitudinal processes, the various authors represented in these pages ponder questions like: Can one harbor both conscious and unconscious attitudes toward the same object? Might such attitudes relate differently to judgments and behavior? What sorts of consequences result from discrepancies between the two attitudes? Also, one would be remiss to overlook the energizing role that recent advances in implicit measurement have played in this research. How might we make sense of discrepancies between implicit and explicit measure of attitudes, and how might the different measures map onto the processes operating within the mind?

We argue that, despite its predating the surge in research on implicit measures, the MODE model provides a cogent and compelling account of many of the attitudinal phenomena that recently have been the focus of attention. These include conceptual issues currently couched in conscious and unconscious terms, as well as inferences premised on dissociations between implicit and explicit measures. The chapter is organized in three major sections. We first will provide a brief review of the MODE model. Then, we will summarize a lengthy series of investigations we have conducted in which implicit measures have been used to test the MODE model, largely in the domain of racial prejudice. In the third and final section, we will broaden the discussion from the initial focus on prejudice to more general issues that have arisen regarding implicit measures. In particular, we will consider the implications of both the MODE model and the associated empirical findings for the interpretation of observed dissociations between implicit and explicit measures of attitude. In so doing, we hope to illuminate how contemporary research findings are consistent with the model.

The MODE Model

Before describing the MODE model, some definitional clarity is in order, particularly with respect to the term *attitude*. It has seen varied definitions over the years (Eagly & Chaiken, 1993), but all of them describe some way in which positivity or negativity is linked to some attitude object. The MODE model identifies this link explicitly by defining attitude as an association in memory between an object and one's evaluation of it (see Fazio, 2007, for an extensive recent discussion of this definition). The strength of this object-evaluation association, as we will see, has some important implications for attitude-behavior processes.

Every attitude can be located somewhere along a strength dimension. Relative to weak attitudes, strong attitudes are stable, are resistant to persuasive appeals, and more reliably predict behavior (Petty & Krosnick, 1995). The MODE model's definition of attitude captures this critical dimension via the object-evaluation association. The weak end of the continuum is marked by the nonattitude, where there is simply no object-evaluation association (such as one's evaluation of the window pane on the left side of one's office window). When asked about one's attitude toward such objects, a respondent must construct an attitudinal response on the spot, even if that were to mean nothing more than

shrugging one's shoulders and claiming perfect neutrality toward the object. As we move along this strength dimension, one finds attitudes with more accessible object-evaluation associations. Such attitudes can be sufficiently accessible that the mere perception of the attitude object automatically and inescapably evokes an evaluative response (such as when the sight of a cigarette immediately prompts cravings in a smoker; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). An abundance of empirical evidence attests to the pervasiveness of this automatic activation effect, as well as its downstream consequences for attitude-behavior processes (see Fazio, 2001 for a review). Indeed, it is with basic cognitive phenomena of attention, perception, and categorization that a relatively spontaneous attitude-behavior process begins.

Spontaneous Attitude-Behavior Processes

The MODE model distinguishes two basic classes of attitude-to-behavior processes. The difference centers on the extent to which pursuing a particular course of action involves a spontaneous reaction to one's perception of the immediate situation versus deliberation regarding the behavioral alternatives. Details regarding the model and relevant empirical findings are available in earlier chapters (Fazio, 1990; Fazio & Towles-Schwen, 1999). Briefly, however, the model postulates that attitudes can guide behavior in a spontaneous manner, without the individual actively considering the relevant attitude and without the individual's necessary awareness of its influence. Instead, the attitude may be activated from memory automatically upon the individual's encountering the attitude object. The automatically activated attitude will influence how the person construes the object in the immediate situation: either directly, as when the activated evaluation forms the immediate appraisal (e.g., an immediate "Yuk" reaction), or indirectly, as when the activated attitude biases perceptions of the qualities exhibited by the object. Ultimately, this construal will affect the person's behavioral response.

Thus, the model postulates that for attitudes strong enough to be automatically activated upon perception of the object, attitude-relevant behavior can flow spontaneously from the attitude, unimpeded by more controlled processes. Clearly, the reasoning is premised on the notion of constructive social cognition first championed by the "New Look" movement (Bruner, 1957) and followed by several decades' worth of

fascinating research illuminating the effects of temporarily and chronically accessible constructs on perception, judgments, and behavior (see Higgins, 1996, for a review). Over the years, our research program has documented many fundamental consequences of accessible attitudes that contribute to this spontaneous attitude-behavior relation. Consistent with this theme, we have demonstrated that accessible attitudes can orient attention (Roskos-Ewoldsen & Fazio, 1992), influence categorization of the attitude object (Smith, Fazio, & Cejka, 1996; Fazio & Dunton, 1997), and bias visual perceptions of the object (Fazio, Ledbetter, & Towles-Schwen, 2000). In addition, they have the potential to influence the processing of information related to the attitude object. That is, the more accessible the attitude, the more likely it is that new information about the object will be disambiguated in an attitudinally congruent manner (e.g., Fazio & Williams, 1986; Houston & Fazio, 1989; Schuette & Fazio, 1995).

This sometimes unwitting tendency to attend to attitude-congruent aspects of an object and to twist its ambiguous qualities into alignment with our attitudes is central to the MODE model's postulate regarding a spontaneous attitude-to-behavior process. By shaping construals of the object in the immediate situation, automatically activated attitudes can influence behavior without any necessary reflection on our parts and without any necessary awareness of the biasing influence of our attitudes. It is important to reiterate, however, that attitude accessibility exerts a critical moderating role. Any such spontaneous process is predicated upon automatic activation of the attitude upon encountering the attitude object. In fact, considerable evidence has accumulated indicating that the extent of biased information processing about an object varies as a function of attitude accessibility (see Fazio, 1995, for a review). For example, in a study of the 1984 presidential election, Fazio and Williams (1986) found that respondents' attitudes toward the candidates predicted impressions of their debate performance, but this relationship grew stronger as attitude accessibility increased.

Deliberate Attitude-Behavior Processes

As apparent as the automatic ways in which attitudes can steer behavior may be, it is also clear that much of our behavior is more thoughtfully determined. That is, instead of a "top-down," attitude-driven process, oftentimes a "bottom-up" process of scrutiny and deliberation precedes a behavioral response. In this latter case, behavior toward a given object is influenced less by the evaluation it may automatically evoke and more by a deliberative comparison of the behavioral alternatives. Under some circumstances, individuals analyze the costs and benefits of a particular behavior and, in so doing, deliberately reflect upon the attitudes relevant to the behavioral decision so as to arrive at a behavior plan, which they may then choose to enact (see Ajzen, 1991; Ajzen & Fishbein, 1980, 2005).

In contrast to the more spontaneous process we described earlier, what is central to the deliberative mode is the exertion of effort. Indeed, it is clear that humans can be as reflective as they are impulsive and that sometimes they are a little of both (Strack & Deutsch, 2004). The critical questions, and those most central to the MODE model, are under what conditions one or the other orientation predominates, and how these processes interact on the path from attitudes to behavior.

The Role of Motivation and Opportunity

Given the effortful reflection required by the deliberative alternative, some motivating force is necessary to induce individuals to engage in the reasoning. The MODE model posits that a variety of motivational factors might push an individual toward a more deliberative scrutiny of behavioral options. Perhaps the most fundamental of these motives is the desire to be accurate: that is, to reach valid conclusions. The MODE model is certainly not alone in arguing for the importance of accuracy motivation. A good example is Kruglanski's (1989) theory of lay epistemics, which identifies "fear of invalidity" as a motive that drives individuals to seek out and consider relevant knowledge. Contemporary models of persuasion such as the Heuristic-Systematic Model (HSM; Chen & Chaiken, 1999) and Elaboration Likelihood Model (ELM; Petty & Wegener, 1999) also assume an accuracy motive. However, other motives—such as the need to belong (Baumeister & Leary, 1995) and to feel positively toward the self (Sedikides & Strube, 1997)—may be similarly capable of pushing the perceiver in a more deliberative direction. But the goals entailed in these motives aren't to arrive at a more accurate conclusion in general, but to reach a more specific and desired conclusion. A motive to belong, for example, might lead one to arrive at a behavioral decision to ingratiate others or to highlight their similarities. As we shall see later, a motivation to avoid prejudice against a given

group may lend itself to more positive behavioral responses to its members. The critical similarity among these various motives, however, is that they all imply the exertion of effort for the purpose of reaching (or avoiding reaching) a given conclusion, whether to be right, liked, accepted, or something else.

Of course, motives do not make behaviors. In order for any motivation to overcome the influence of one's attitude, an opportunity for this motive to exert its influence must also be available. And again, the model views "opportunity" broadly; it manifests in a variety of ways. Opportunity can simply amount to a matter of time; careful consideration of information simply cannot be done quickly (e.g., Jamieson & Zanna, 1989). But opportunity also comes in psychological forms. Because our cognitive resources are limited, fatigue, distraction, and other factors can interfere with one's ability to process information. For example, Baumeister and colleagues' work on self-regulation suggests that "ego-depletion" might compromise one's processing resources, thus limiting the potential impact one's motivated intentions can have on judgments and behaviors (Baumeister, Bratlavasky, Muraven, & Tice, 1998).

According to the MODE model, these two moderating factors—motivation and opportunity—determine the extent to which the attitude-to-behavior process is primarily spontaneous versus deliberative in nature. Moreover, as implied by the preceding analysis and illustrated by subsequent research, a shift toward a more deliberate mode of processing requires both motivation and opportunity. For example, Sanbonmatsu and Fazio (1990) found that when participants were deciding between two alternatives, they engaged in the effort of retrieving specific attribute information from memory (beliefs about the alternatives) only under conditions in which motivation to reach a valid decision was heightened and no time pressure was applied. When their fear of invalidity was relatively low or when their decision-making was done under time pressure, participants showed evidence of relying on their global attitudes instead of specific attribute knowledge (see Fabrigar, Petty, Smith, & Crites, 2006, for related evidence).

In addition to delineating two distinct classes of attitude-behavior processes, the MODE model explicitly postulates the possibility of processes that are neither purely spontaneous nor purely deliberative, but instead are what we refer to as *mixed* processes, ones that involve a combination of automatic and controlled components. Any controlled component within a mixed sequence requires, once again, that the individual both be motivated to engage in the necessary cognitive effort

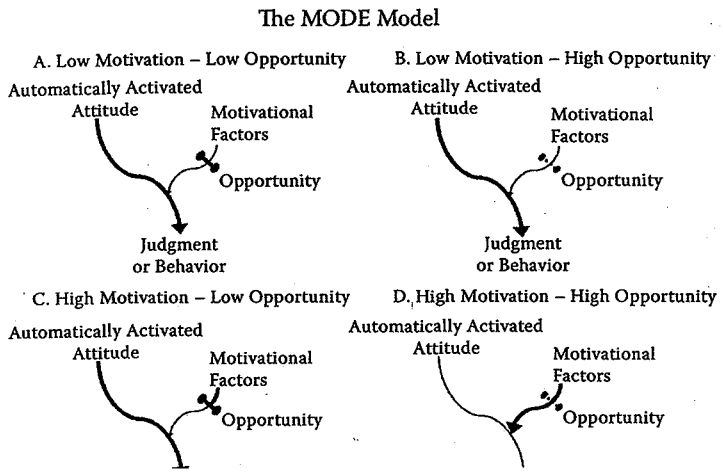

FIGURE 2.1 Depictions of the MODE model as a function of high versus low motivation and opportunity.

and have the opportunity to do so. Figure 2.1 provides a graphical summary of the MODE model, illustrating the potential interplay between automatic and controlled processes. The model views an automatically activated attitude as a "starting point" for judgment and behavior. However, the "downstream" consequences of the automatically activated attitude—that is, its influence on overt judgments—can be moderated by motivation and opportunity. In the figure, the thickness of a "stream" is intended to convey the extent of influence. As depicted in the top two panels, when little or no motivation to deliberate is evoked in a given situation, overt judgments and behavior are hypothesized to reflect the direct influence of the automatically activated attitude. These downstream consequences of attitude activation would occur via the mechanisms discussed earlier as comprising a spontaneous attitude-to-behavior process. Any motivational factors that might be evoked can exert an influence on the overtly expressed judgment, assuming that the situation and behavior in question provide sufficient opportunity for controlled, deliberative processing. Essentially, the opportunity factor can be viewed as a gating mechanism that determines the extent to which motivational factors can influence the overt judgment. When the gate is open, as in panel D of the figure, motivational goals can have a strong influence on their overt judgments, potentially attenuating the influence of the automatically activated attitude. However, such motivated efforts will be thwarted when the opportunity to deliberate is minimal, as in panel C.

One of the first experiments to examine such mixed processes was conducted by Schuette and Fazio (1995), who tested the hypothesis that attitudinally biased information processing should be jointly affected by attitude accessibility and motivation to deliberate. Schuette and Fazio employed the paradigm developed by Lord, Ross, and Lepper (1979), in which participants' judgments of the quality of presumed scientific studies concerning the deterrent efficacy of capital punishment are potentially biased by their own attitudes toward the death penalty. On the basis of the MODE model, Schuette and Fazio reasoned that any such biasing effects should depend on the accessibility of those attitudes and the presence of motivation. In some conditions, the accessibility of participants' attitudes toward the death penalty was enhanced by having them repeatedly express their attitudes in an early phase of the experiment. Such repeated expression is known to increase the likelihood of automatic attitude activation when the attitude object is later encountered (Fazio et al., 1986). Replicating earlier findings (Houston & Fazio, 1989), participants in the repeated-expression condition displayed greater attitudinally biased processing than did those in a condition in which attitudes had been earlier expressed only a single time. That is, the correlation between attitudes and judgments of the research quality was stronger for individuals with more accessible attitudes. However, the MODE model predicts that such attitudinal bias should be minimized in the presence of motivation and opportunity. In this study, participants in all conditions were provided ample opportunity to digest the information, but some participants were particularly motivated to process the information carefully. These individuals were told that their responses would be made public and compared to the judgments rendered by a panel of experts. No such fear of invalidity was invoked for participants in the low-motivation conditions, and it was within these conditions that those with more accessible attitudes exhibited biased processing. They judged the quality of the research in accordance with their attitudes, agreeing with research supporting their views and criticizing research that contradicted them. (Note that this finding accords with panel B of Figure 2.1.) The more motivated participants, on the other hand, were able to overcome the biasing effects of their attitudes and presumably judge the research more objectively, even when their attitudes were highly accessible. (This latter finding matches the predictions depicted in panel D.)

This interplay between the biasing effects of automatically activated attitudes and "corrective" measures prompted by a relevant motivational factor lies at the heart of the MODE model. Automatically activated attitudes can have a potent effect on overt judgments and behavior, but their influence can be attenuated when some relevant motivational goal arises. In the Schuette and Fazio (1995) research, both motivation and the attitude's capacity for automatic activation were manipulated experimentally. In subsequent work, we adopted an individual difference approach to testing the MODE model. Automatically activated attitudes were assessed via an implicit measure. In addition, individuals who experienced varying levels of motivation to counter the influence of their automatically activated attitudes were identified. Such an approach required that there was variability across individuals with respect to both their automatically activated attitudes and relevant motivations. The domain of racial prejudice—a richly active area of social psychological research—proved very well suited to these research aims. We now have accumulated a series of empirical findings confirming the predicted moderating role of motivation on the relation between automatically activated racial attitudes and various race-related judgments. Essentially, the estimates of racial attitude provided by an implicit measure are predictive for individuals low in motivation to control prejudiced reactions, but the relation is attenuated, and often reversed (a pattern indicative of motivated overcorrection), as motivation increases.

MODE Model Applications to Racial Prejudice

Most social psychological treatments of prejudice entail both automatic and controlled components. Devine (1989), for example, argued that the automatic component of prejudice is acquired through passive socialization processes, and that nonprejudiced individuals are marked by a controlled, value-driven system to avoid allowing automatic prejudices from influencing their behavior. Most "modern" theories of racism hint at the dual interplay of automatic and controlled processes, but the theory of aversive racism posits explicitly that aversive racists tend to be prejudiced at more automatic (and perhaps less conscious) levels, but still think of themselves as egalitarian (Dovidio & Gaertner, 1998). We argue that the MODE model provides a broad, overarching means of conceptualizing and explaining the various roles that automatic and controlled processes play in such discriminatory behavior. Because many individuals wish to avoid prejudiced responses (or at least their appearance), motivational forces should interact with automatically activated racial attitudes in pre-

dicting race-related behavior—at least when opportunity allows. Because we view attitudes as the starting point of race-related behavior, we first discuss how we have conceptualized and assessed racial attitudes, particularly their automatic properties.

Automatically Activated Racial Attitudes

The priming measure. The research involves not only the application of the MODE model to the domain of racial attitudes, but also our technique for assessing automatic attitude activation. The studies concern the direct assessment of the evaluations that are automatically activated in response to Blacks. They employ a priming procedure that was first developed in the mid-1980s (Fazio et al., 1986) and has since been used widely to study automatic attitude activation (see Fazio, 2001, for a review). Briefly, the participants' task on each trial is to indicate the connotation of an adjective as quickly as possible: Does it mean "good" or "bad"? We are concerned with the latency with which this judgment is made and, more specifically, the extent to which responding is facilitated by the prior presentation of a prime. The pattern of facilitation that is exhibited on positive versus negative adjectives provides an indication of the individual's attitude toward the primed object. Relatively more facilitation on positive adjectives is indicative of a more positive attitude and relatively more facilitation on negative adjectives is indicative of a negative attitude. Furthermore, these estimates are obtained without the individual's awareness that his or her attitude is even being assessed; the participant is not asked to consider his or her attitude toward the prime during the task. Yet, it is possible to infer from the facilitation data the degree to which positive or negative evaluations are activated when the object is presented.

We have applied this methodology to the assessment of racial attitudes. We will provide only a brief sketch of the procedure here; details are available in Fazio, Jackson, Dunton, and Williams (1995). Participants are told that the experiment concerns word meaning as an automatic skill and that a variety of tasks will be performed. The procedure consists of four phases, the last being the actual priming task. The purpose of the first task is to obtain baseline latency data. On each trial, the participant is presented with an adjective (e.g., *attractive, likable, disgusting, offensive*) and asked to indicate as quickly as possible whether it means "good" or "bad." The next two phases are intended to prepare participants for the priming task, which involves the presentation of faces as primes and adjectives as targets. The second phase is

presented to the participants as involving the ability to learn faces. They simply attend to a series of head-shots of individuals presented on the computer screen. The third phase is a recognition test. Participants are presented with a face and asked to indicate whether the face was one that they had or had not seen in the previous task. Next, the actual priming task occurs. Participants are told that the previous tasks will now be combined, in the interest of determining the degree to which judging word meaning is an automatic skill. The experimenter indicates that if it truly is, individuals should be able to perform just as well as in the very first phase of the experiment, even if they have to do something else at the same time. In this case, the task to be performed simultaneously is learning faces. Thus, this phase of the experiment is said to involve both the learning of faces and the judgment of adjectives. On the target trials, images of White and Black students serve as the primes, again followed by positive and negative adjectives. We record the latency to respond to the adjectives as a function of prime race.

This procedure yields a multitude of observations for each participant. We routinely have reduced the data from any given respondent to a single index that serves as the estimate of the individual's attitude toward Blacks. To do so, average facilitation scores are computed for each person on positive and negative adjectives for each face that was presented. This preliminary step yields mean facilitation scores for each of the multiple White and Black faces. Thus, it is possible to examine the interaction of race of photo × valence of adjective for each participant. The effect size of this interaction is computed and serves as our estimate of the individual's attitude. Given the computational procedure, more negative scores reflect a pattern of facilitation indicating greater negativity toward Blacks: relatively more facilitation on negative adjectives when they were preceded by a Black face than a White face and relatively less facilitation on positive adjectives when they were preceded by a Black face. The opposite pattern yields a positive score.

Predictive validity of the priming measure. In adopting this paradigm to the study of racial attitudes, we were questioning the assumption that evaluative responses to Blacks reflected the cultural stereotype and, hence, were universally negative (Devine, 1989). Instead, we advocated that meaningful individual differences in White individuals' automatic evaluative responses might be observed. Indeed, across many studies to date, wide variability has been found across hundreds of participants, with many individuals exhibiting various degrees of automatically activated negativity toward Blacks, but others exhibiting more positive automatic responses.

This variability we observe is meaningful. Given the MODE model's emphasis on the "early" influences of automatically activated attitudes, these attitudes should operate subtly, guiding attention, defining situations, and often guiding behavior in a relatively spontaneous fashion. Consistent with this reasoning, attitude indices derived from the priming measure have proven predictive of a race-related behavior across a number of studies. For example, in one such study, participants were "debriefed" by a Black experimenter after completing the priming measure. After a several-minute interaction with a given participant, the experimenter rated the extent to which the participant seemed interested and friendly, paying particular attention to nonverbal behavior such as smiling, eye contact, and distance. These experimenter ratings showed clear correspondence with the attitude index—participants characterized by negativity toward Blacks produced a less friendly and interested impression in the eyes of the Black experimenter (Fazio et al., 1995).

Additional evidence of construals being shaped by automatically activated racial attitudes was found in a study in which White participants were charged with rating the quality of essays purportedly written by a Black undergraduate (Jackson, 1997). After completing the priming measure, participants returned for a second session, where they were told that they would be serving as judges for an essay contest. After reviewing biographical information about the author that revealed his race, participants rated how interesting, well-written, and persuasive the essay was, as well as how deserving it was to win the contest. A composite index of participants' ratings was again correlated with their attitude estimates: negative automatically activated racial attitudes appeared to guide participants' impressions of the work created by a Black student.

We recently reported similar findings in the context of a study about committee selection procedures (Olson & Fazio, 2007a). As in Jackson's research, participants had completed the priming measure prior to returning to the lab for a "second study." Upon returning, they were told that they would be simulating the work of committee members, who must review a large number of applications in a limited amount of time. Applications of four individuals who were applying to work in the Peace Corps were provided for them to review. The first two applicants consisted of highly qualified and poorly qualified White females. The critical applicants were the final two: moderately qualified Black and White males. After reviewing the extensive application materials (including school transcripts, work history, personal statement, and an interview summary purportedly provided by a Peace Corps official),

they rated each applicant on a number of dimensions including their credentials, likeability, and suitability for Peace Corps work. These ratings were averaged and a relative preference score for the Black versus the White critical applicants was computed for each participant. As expected, the attitude index predicted their ratings: negativity toward Blacks according to the priming measure corresponded to a preference for the White relative to the Black applicant.

Comparable findings can be found in research employing variations of the priming measure. For example, Dovidio, Kawakami, Johnson, Johnson, and Howard (1997) reported correspondence between the nonverbal behaviors exhibited by White participants when interacting with a Black experimenter and a priming measure using composite sketches of Blacks and Whites as primes. Individuals characterized by negative attitude indices showed reduced eye contact, among other things, when interacting with a Black relative to a White (see also Dovidio, Kawakami, & Gaertner, 2002).

In addition to attesting to the predictive validity of the priming measure, these studies also validate the role of attitude accessibility in the attitude-behavior process. Instead of manipulating or measuring attitude accessibility and then examining how it moderated the relation between self-reported attitudes and behavior, as in early MODE model research, the more recent work essentially incorporated accessibility into the very measure of attitude itself. That is, the priming measure assessed the evaluation automatically activated by the presentation of African-American faces, and the resulting attitude estimates proved predictive of race-related judgments and behavior. Thus, these findings provide support for the MODE model's postulate that behavior can be a direct reflection of the attitude automatically activated from memory upon exposure to the target.

But this, of course, is only part of the story. Despite the correspondence we have observed between automatically activated attitudes and race-related judgments and behavior, oftentimes we see no evidence for that simple relationship. The MODE model predicts that additional processes, afforded by motivation and opportunity, can be evoked to steer behavior somewhere other than where the automatically activated attitude would imply. We turn to these motivational factors next.

Motivation to Control Prejudiced Reactions

According to the MODE model, deliberative processes enter into the attitude-behavior relation given the presence of some relevant moti-

vation. Earlier we described some of the motives (e.g., accuracy) that operate in a variety of attitude domains. Although they probably operate in the prejudice domain as well, additional motivational forces are uniquely at work when it comes to prejudice and make it all the more useful as an arena for testing the MODE model. Specifically, many White individuals wish not to appear prejudiced, either to themselves or to others. Others are more motivated to avoid conflict or dispute with respect to race. We have employed our Motivation to Control Prejudiced Reactions (MCPR) scale as a means of assessing these motives. The MCPR consists of two orthogonal factors (Dunton & Fazio, 1997). The first, concern with acting prejudiced, reflects a personal commitment to avoid reactions that others and oneself might consider prejudiced, and contains items such as, "I feel guilty when I have a negative thought or feeling about a Black person." The second factor, restraint to avoid dispute, involves a willingness to inhibit the expression of one's own thoughts and feelings in the interest of avoiding dispute with or about Blacks, and contains items like, "If I were participating in a class discussion and a Black student expressed an opinion with which I disagreed, I would be hesitant to express my viewpoint." Importantly, neither factor correlates with the priming estimates of automatically activated racial attitudes.

Before turning to research examining the moderating influence of these motivational factors, we'll review some of what we have learned about the two factors. First of all, they have very different correlates. A correlation of .50 was observed between scores on the concern factor and the endorsement of egalitarian values, as assessed by Katz and Hass's (1988) Humanitarianism-Egalitarianism Scale (see Fazio & Hilden, 2001, footnote 1). Restraint factor scores, on the other hand, were independent of egalitarianism, $r = -.01$. Additional correlates were identified in a study in which some 150 college students completed the MCPR during a mass survey, the priming procedure in an initial lab session, and a survey regarding their race-related childhood experiences in yet another session (Towles-Schwen & Fazio, 2001). A number of intriguing relations emerged. For example, the attitude estimates correlated significantly with the positivity of relatively recent interactions with Blacks, those during the high school years. What may have been most interesting about the study, however, was that the two factors of the MCPR related in very different ways to the past experience variables. Higher scores on the concern factor were associated with reports of more positive interactions with Blacks at all school levels and reports of little parental prejudice. Scores on the restraint factor correlated with

these same variables, but in the reverse direction. Greater restraint was associated with relatively infrequent interactions, which (when they occurred) were less positive, and with relatively high parental prejudice. Thus, higher restraint seems to be associated with a lack of contact with, and possible avoidance of, African Americans.

Further evidence that high-restraint individuals adopt an avoidance strategy comes from a study in which White individuals provided private evaluations of Black and White job applicants prior to making videotaped public statements evaluating their credentials (Olson & Fazio, 2007b). Not surprisingly, according to naïve judges who rated the transcripts of the White participants' spoken words, participants' public statements corresponded with their private evaluations of the job applicants. However, this relation was moderated by scores on the restraint factor. When discussing a Black candidate (relative to a similarly qualified White target), high-restraint individuals showed less correspondence between their privately reported beliefs about those individuals and the words they chose to utter publicly about them. It appears, then, that high-restraint individuals obfuscate their private views when speaking publicly about Blacks, perhaps in an attempt to avoid controversy or dispute. The concern factor showed no such pattern.

We also have examined how the two motivational factors, as well as automatically activated racial attitudes, relate to individuals' emotional reactions following their exhibiting a seemingly prejudiced response (Fazio & Hilden, 2001). In a study about "emotional reactions to television commercials," participants were exposed to a series of Clio Award-winning ads, including the target public service ad. This ad induces viewers to assume wrongly that the African American who is pictured and the criminal who is described in the scrolling text next to the image are one and the same individual. Instead, viewers eventually learn that the photo is actually of the police officer who apprehended the criminal. Emotional reactions to the ad varied as a function of automatically activated racial attitudes and the two MCPR factors. More positive attitudes were associated with feelings of guilt. Individuals characterized by higher scores on the concern with acting prejudiced factor reported both greater guilt and greater agitation. Those with higher scores on the restraint to avoid dispute factor experienced agitation but not guilt. Thus, the unique experience of guilt (unaccompanied by any other feelings of agitation) was associated with positive racial attitudes, ones that were so well-internalized that they were capable of automatic activation. Guilt and agitation were more pronounced among those for whom the ad provoked violation of a valued "ought" standard (Higgins, 1987)

regarding egalitarianism (higher concern), whereas agitation alone was accentuated among those for whom the ad was reminiscent of the very race-related dispute they seek to avoid (higher restraint).

Mixed Processes in the Racial Prejudice Domain

We turn now to the evidence for mixed processes in racial prejudice. In the research we review in this section, participants typically completed both the priming measure of racial prejudice and the MCPR prior to performing some race-related judgment or behavior. Our analytic approach has been to regress these judgments or behaviors onto the attitude estimate provided by the priming measure, scores from the MCPR factors, and importantly, their interaction terms. Mixed processes are revealed in interaction effects, whereby the direct effect of the attitude estimate is reduced (or even reversed) as motivation increases. Throughout these studies we have found consistent evidence for the MODE model's predictions regarding these mixed processes. However, these interaction effects can take varied forms, which, as we shall see, reveals interesting insights about the correctional goals involved in motivation to control prejudiced reactions.

Moderating effects of the concern factor. Some of the first evidence of these mixed processes in the racial prejudice arena was found with respect to participants' responses on what has been indisputably the most popular self-report measure of racial prejudice: the Modern Racism Scale (MRS, McConahay, 1986). This seven-item measure prompts respondents to indicate their agreement with statements like "Over the past few years, Blacks have gotten more economically than they deserve," and "Blacks should not push themselves where they are not wanted."

It is important to note that indicating responses to such statements is itself race-related verbal behavior. From the MODE model's perspective, such verbal expressions occur further "downstream" in the attitude-behavior process, after any automatic activation of racial attitudes, and after any motivation to control prejudiced reactions has been evoked. Thus, responses to these statements—and any explicit measure of prejudice—have the potential to be influenced by both automatic and more motivated forces. Indeed, evidence suggests that the MRS, contrary to how it was originally portrayed, is a reactive measure; White participants report less prejudice on it in the presence of a Black experimenter (Fazio et al., 1995; see Olson, in press, for a more extensive review). The MODE model predicts that in the absence of motivation, automatically activated racial attitudes should directly influence responses on the MRS.

However, motivated individuals should wish to avoid the influence of any automatic racial prejudice and respond differently to the items.

This is just what we found (Dunton & Fazio, 1997). Individuals who were relatively unconcerned about acting prejudiced responded to the MRS items in a manner that was consistent with the attitude index provided by the priming measure. Those with negative automatically activated racial attitudes reported more prejudicial beliefs on the MRS, and those with relatively positive racial attitudes reported less prejudice. However, as concern with acting prejudiced increased, correspondence between the two measures decreased, so much so, in fact, that participants characterized by more negative attitudes and higher scores on the concern factor seemed to have gone out of their way to avoid the appearance of prejudice. They responded even more positively to the MRS items than did participants who displayed positivity in response to photos of African Americans during the priming procedure. (Although we will be discussing Figure 2.2 extensively in a subsequent section, readers who wish to see a graphical depiction of the form of the obtained interaction will find it represented in panel B of the figure.)

A similar moderating effect of the concern factor was observed in a study concerning people's expressed willingness to enter situations

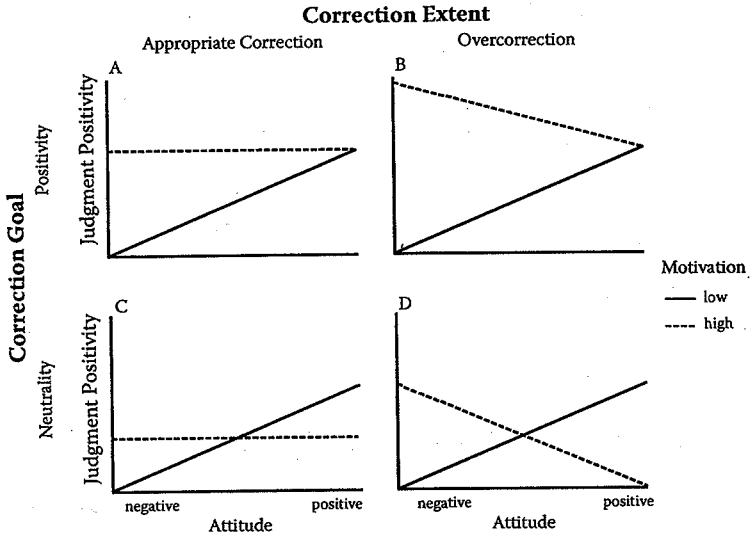

FIGURE 2.2 Hypothetical forms of attitude × motivation interactions predicting race-related judgments and behaviors. Reprinted from Olson & Fazio (2004).

involving a Black interaction partner (Towles-Schwen & Fazio, 2003). As Snyder and Gangestad (1982) emphasized, the social situations we choose to enter, or avoid entering, can be a reflection of the social worlds we wish to create. Towles-Schwen and Fazio speculated that the choice to enter or avoid certain social situations involving Blacks might be determined by the mixed processes posited by the MODE model. Participants were invited to imagine themselves about to enter a variety of social situations (e.g., granting a maintenance worker access to one's home, sharing a small dorm room, sitting down at a crowded table where someone is already seated). They were asked to rate how comfortable they would be to enter these situations. After responding to these situations in the absence of any specification of the interaction partner, participants rated their comfort with the situations again, this time imagining each of a variety of hypothetical interaction partners, one of whom happened to be Black. Interestingly, a marked difference was observed when participants imagined a Black partner relative to their ratings when the partner was left unspecified. In general, participants claimed that they would find the social interaction more comfortable when a Black partner was involved. However, the extent to which they did so varied as a function of automatically activated attitudes and concern with acting prejudiced. Consistent with the MODE model, more positive attitude estimates corresponded to greater anticipated comfort interacting with a Black partner, but only among participants with relatively low scores on the concern factor. Participants who were more concerned about acting prejudiced tended to show a reverse pattern. Those claiming the most comfort were those marked by negative automatically activated attitudes and high concern (again, see panel B of Figure 2.2). It appears that these individuals were overcompensating for their prejudice when indicating their willingness to pursue an interaction with a Black partner.

Moderating effects of the restraint factor. Moderating effects of motivation on the relation between automatically activated racial attitudes and a race-related judgment have been observed in additional studies. In the two we wish to highlight now, the driving motivational force proved to be restraint to avoid dispute. Dunton and Fazio (1997) asked participants to list the feelings that came to mind when they thought of the "typical Black male undergraduate." They then rated their own thoughts in terms of their positivity and negativity, which were averaged to form an overall index of their self-reported evaluations. In later sessions, participants completed the priming measure of racial attitudes and the MCPR. The regression analysis revealed a significant interaction

between attitudes and the restraint to avoid dispute factor of the motivation scale. Correspondence between estimates of automatically activated racial attitudes and participants' ratings of their feelings toward the typical Black male undergraduate increased as restraint decreased. Those with low restraint scores appeared to be guided simply by their attitudes. More motivated participants, on the other hand, showed a reverse pattern, indicative of correction for their automatically activated attitudes. (The form that the interaction assumed is represented in panel D of Figure 2.2, which will be discussed shortly.)

We observed a conceptually parallel finding in a study that we conducted concerning "first impressions" (Olson & Fazio, 2004). We asked participants to offer their first impressions of a series of individuals depicted in photographs that were presented on their computer screens. These were images of various people—men, women, Blacks, Whites, and people of other races—in various occupational settings. Included therein were several Black-White pairs, matched in terms of the status and independence of their occupational roles (e.g., a Black male potter and a White male bricklayer; a Black male minister and a White male professor). Participants recorded their impressions of each target individually on a variety of scales (e.g., likeability, competence) as they appeared on the screen. The analyses focused on the difference in the ratings of the Black target persons relative to the matched White targets. In general, participants reported viewing the Black targets as more positive than the White targets. However, we also observed an interaction indicating that this preference for Blacks was more characteristic of some kinds of individuals than others. Among individuals with low scores on the restraint to avoid dispute factor, racial attitudes corresponded with the trait inferences. Those with more positive attitudes judged the Black targets more positively than the White targets. This relation was attenuated and, once again, even reversed, as restraint to avoid dispute increased (see panel D of Figure 2.2). Thus, a correction process appeared to be at work among the more motivated participants.

Correctional goals implied by the motivational factors. We now have seen repeated instances of motivation to control prejudiced reactions moderating the relation between automatically activated racial attitudes and some race-related judgment or behavior. In some cases, concern for acting prejudiced seemed to do the work, and in other cases it was restraint to avoid dispute. For example, it was the concern factor that interacted with attitude estimates in predicting MRS scores (Dunton & Fazio, 1997) and reported comfort in entering social situations with a Black individual (Towles-Schwen & Fazio, 2003). Restraint, on

the other hand, played a moderating role in predicting evaluations of a Black male undergraduate (Dunton & Fazio) and first impressions of Black individuals (Olson & Fazio, 2004).

We have hesitated to make predictions about the sorts of social judgments and behaviors that might be more likely to evoke one or the other motivational factor. Looking back across several studies, however, some patterns have emerged that we believe provide some insight as to the sorts of situations that lend themselves to the influence of the two motivational factors, as well as the sorts of corrective goals each motivation implies. Our speculations hinge on whether the Black social target in question is construed at the category or the individual level. Take, for example, the items included in the MRS. They prompt respondents to consider Blacks as a group, at the category level. In Towles-Schwen and Fazio's (2003) work, the social target with whom participants were imagining interacting was described simply as Black, and described only at the category level. In research involving restraint factor interactions, on the other hand, the Black social targets were depicted more as individuals (albeit as members of a category). So while the concern factor's focus appears to be Blacks as a group, restraint works at the individual level.

The relevance of this individual-versus-group distinction becomes apparent when considering the underlying bases of each of the motivational factors. Recall the concern factor's relatively strong correlation with egalitarianism. If one's impetus to avoid prejudice is based on the institutional factors implicated in egalitarian beliefs (such as a belief in Blacks' historical plight and continued educational and economic disadvantages), then one is more likely to see prejudice at the group level and aim one's own responses to it commensurately. Restraint, on the other hand, is premised on the desire to avoid dispute, conflict, and confrontation. With few exceptions, an individual comes into conflict not with categories of people, but with individuals. If it is the fear of being accused of prejudice, or of being offensive in some way, that characterizes people high in restraint motivation, then such motivation is more likely to be evoked in the interpersonal situations, where conflict actually occurs. So while speculative, we believe that concern motivation is more aligned with the possibility of prejudice toward the group, whereas restraint motivation stems from desire to avoid dispute at the individual level.

Another difference that seems to emerge between the concern and the restraint factors is the pattern of interactions we have observed involving attitude estimates and the motivational factor. As we noted parenthetically earlier, the interactions involving concern and those involving restraint have assumed somewhat different forms. Although admittedly speculative, consideration of these differential patterns provides some further insights regarding the specific correctional goals involved with each motive. The analysis we have offered (Olson & Fazio, 2004) rests heavily on Wegener and Petty's (1995) flexible correction model, which suggests that upon suspicion that they are falling prey to an undesired judgmental bias, individuals may attempt to correct for the bias. They may adjust their judgments on the basis of naïve theories that they hold regarding the direction and magnitude of the unwanted influence.

Hypothetically, the attitude by motivation interaction implying motivated correction can assume one of several potential forms. These are illustrated in Figure 2.2. Common to all the panels of the figure, correspondence between automatically activated racial attitudes and judgments is evident at low levels of motivation, in accordance with the MODE model. The slopes and intercepts of the high-motivation regression lines, on the other hand, vary. The top two panels (A and B) depict correction for negativity only. In panel A, the line is flat, suggesting appropriate correction for negativity. Panel B also implies correction for the purpose of avoiding negative judgments, but in this case, judgments of Blacks among prejudiced individuals are particularly positive—more positive than even low-prejudiced participants. That is, overcorrection is apparent. Findings involving interactions between attitude estimates and the concern factor have shared a remarkably similar form, that of panel B. Here, participants appear to have corrected for automatically activated negativity toward Blacks, but not for positivity.

Panels C and D, on the other hand, depict correction for both negativity and positivity. For example, in panel C respondents with negative attitudes toward Blacks have adjusted their responses in a positive direction, and those with positive attitudes have adjusted their responses in a negative direction, in both cases coming to more closely resemble respondents with neutral attitudes. Panel D also depicts correction for both positive and negative prejudices, but respondents here, as in panel B, appear to have "overshot" their goal, such that the relationship between attitudes and judgments for high-motivation respondents is actually negative. Findings involving interactions between attitude estimates and the restraint factor have also shared a similar form, that of panel D. Here, participants appear to have corrected for both negativity and positivity toward Blacks. It appears, then, that two general correctional goals underlie motivation to control prejudiced reactions,

positivity and neutrality, and these motives align, respectively, with concern and restraint.

Why are the two motives associated with different correctional goals? Given its strong ties to egalitarianism beliefs, as described earlier, we suspect that the historical mistreatment of Blacks is a salient belief among individuals characterized with concern motivation. Their goal is to treat such disadvantaged individuals more favorably. Hence, those motivated by egalitarian values may correct for any negativity that they experience. On the other hand, individuals with more positive attitudes believe they have nothing to correct for; their automatically activated positivity toward Blacks is concordant with their goals.

Restraint motivation, on the other hand, has a more bidirectional quality to it. Recall that the primary motive in the case of restraint is to avoid dispute, not redress historic inequalities. Here our suspicion is that individuals characterized by such motivation may fear accusations of "reverse discrimination" in addition to the more straightforward accusation of prejudice. Indeed, both our research and others' suggest that Whites are often willing to "bend over backwards" to avoid the appearance of prejudice (Biernat & Vescio, 1993; Gaertner & Dovidio, 1977; Jussim, Coleman, & Lerch, 1987; Olson & Fazio, 2004). It's likely, then, that overtly positive treatment of Blacks might sometimes be met with suspicion that such treatment is because of race. Thus, race-related disputes can stem from the appearance of either mistreatment or special treatment of Blacks. High-restraint individuals appear to respond to such dispute-provoking situations by either appearing more positive toward Blacks (if their underlying, automatically activated racial sentiments are negative) or appearing more negative toward Blacks (if their underlying racial sentiments are positive).

Interestingly, a commonality that has emerged across the unidirectional corrective nature of concern and bidirectional corrective nature of restraint is that of overcorrection, as displayed in panels B and D of Figure 2.1. In terms of Wegener and Petty's (1995) flexible correction model, it appears that motivated perceivers believe their automatically activated attitudes to be a potent source of judgmental bias.

Setting the specifics of the correctional process aside for the moment, the studies summarized above have provided support for the MODE model's postulate regarding mixed processes. Verbally expressed judgments—whether they be responses to an explicit measure of racial attitude, first impression ratings, or behavioral intentions regarding a willingness to initiate interaction—all occur farther "downstream" than the automatic activation of any relevant attitude. They can be influenced by automatically activated racial attitudes. But such explicit judgments also may be influenced by motivational factors that can override the effect of the automatically activated attitude.

The Role of Opportunity

In the research summarized above, we have focused on the interactive roles of automatically activated racial attitudes and motivation to control prejudiced reactions. The outcome variables concerned verbally expressed judgments, and these judgments were offered under conditions in which the opportunity parameter of the MODE model was high. That is, it is not at all difficult to monitor and control one's responses to the simple judgmental scales that participants completed. Thus, in all these cases, ample opportunity (e.g., time and cognitive resources) was available for motivated processes to influence race-related judgments.

In the complexity of real-world interracial contact, however, opportunity may be limited. Some of these limits stem from the confines of our cognitive endowments. For example, in contrast to verbal judgments, the nonverbal behavior one emits is less likely to be monitored and, to some extent, is less susceptible to conscious control (DePaulo & Friedman, 1998). Thus, even if one is motivated to control prejudiced reactions, automatically activated negativity might "leak" into the nonverbal channels in interracial settings because there is less opportunity for behavioral control. This is just what Dovidio and colleagues (1997) found in research that employed a priming measure of racial prejudice quite similar to our own. Participants in this research also completed explicit questionnaire measures on racial attitudes prior to interacting with Black and White confederates. Although the explicit measure predicted self-reported liking of the confederates, the priming measure predicted nonverbal expressions like blinking and eye contact (see also Dovidio et al., 2002). Thus, in accordance with the MODE model, motivation is unable to overcome one's automatically activated attitudes when opportunity is not available.

The multiple demands of social life also mean that opportunity is likely to wax and wane. Those with whom we have long-term relationships are likely to see us when we are tired or preoccupied, that is, when our "true colors" are likely to emerge. It is unlikely that Whites in long-term relationships with Blacks are always able to suppress any underlying automatic negativity toward Blacks in these low-opportunity moments. Thus, automatic prejudice is likely to emerge periodically in real interracial relationships that extend over time, and influence the

quality—and longevity—of those relationships, even when the White individual is motivated to control prejudiced reactions. That is, when opportunity is low, as it sometimes is in real relationships, motivation to control prejudice will be impotent to overcome automatic prejudice.

The desire to examine automatically activated attitudes and motivation in the context of real-life relationships prompted Towles-Schwen and Fazio (2006) to conduct a pair of logistically complex field investigations that proved very informative. The studies took advantage of the fact that Indiana University randomly pairs any freshmen who have requested a double dormitory room and not specified a roommate by name. Hence, with the cooperation of the Housing Office, it was possible to identify dyads consisting of an African American and a White freshman who had been randomly paired to share a room. The White member of such dyads was recruited to participate, for monetary payment, in a study presumably concerned with satisfaction with college life. A comparison sample of Whites who had been randomly paired with another White also was recruited. The initial study confirmed the presumption that these interracial relationships are problematic. At the end of the fall semester, Whites paired with an African American reported less satisfaction with their roommates, as well as a lower frequency of engaging in various joint activities. However, the most striking evidence of the difficulties faced by the interracial dyads came from a very simple objective variable: the sheer likelihood of the relationship remaining intact. By the end of the semester, a number of these relationships had simply dissolved; one or the other person had moved out. In fact, 28% of the interracial dyads had split up, a dissolution rate that was significantly higher than the 9% observed for the same-race dyads.

Study 2 focused on whether knowledge of the White partner's automatically activated racial attitudes, as assessed early in the semester, could predict the success of the interracial roommate relationships. Again, the students were recruited for a study presumably concerned with satisfaction with college life. In an initial session, they completed a number of computer tasks and questionnaires, many of which were intended to obscure the interest in racial attitudes and in their roommate relationships. The priming measure of racial attitudes was embedded within a series of speeded computer tasks that presumably assessed cognitive skills potentially related to doing well in college. Participants also completed opinion surveys regarding a wide array of social and political issues, and the MCPR was embedded within those items.

Housing Office records revealed that nearly 30% of the interracial roommate relationships had dissolved by the end of the first semester; 57% of the dyads failed to remain intact for the entire academic year. The number of days the roommates were together, which ranged from 24 to 252 days, provided a very useful continuous measure of the success of the relationship. The measure of automatically activated attitudes correlated significantly with this duration index; the more positive the attitudes, the greater the longevity of the relationship. Importantly, neither factor of the MCPR bore any relation to the duration of the relationship, nor did either of these motivational factors moderate the relation between attitude and status of the relationship. As suggested earlier, this lack of effect for motivation likely stems from the opportunity factor. No matter how well intentioned one might be, no matter how much one may wish to monitor one's behavior carefully, no matter how much one might try to control seemingly prejudiced reactions, it is just not possible to do so in the sort of intimate, continuous interaction context that characterizes sharing a dorm room.

Although all of this research attests to the importance of the impact of the opportunity factor, in our research on racial prejudice we have not pursued experimental manipulations of opportunity to provide a more rigorous test of its moderating role. However, as we shall see later, other researchers have pursued just this in research on "ego-depletion" and related constructs.

Failures to Evoke Relevant Motivations

The MODE model describes opportunity as a factor moderating the influence of automatically activated attitudes on judgments and behavior, and it can vary by person, situation, or judgment/behavior type. We have argued that in cases where verbal judgments are made with unlimited time and no source of distraction, opportunity is quite high and hence should not limit the influence of motivational forces aimed at curbing the influence of automatically activated attitudes. However, in the laboratory's own research, we have seen two cases where motivation failed to influence judgments even under such "ideal" conditions. In Jackson's (1997) research, where participants rated the quality of an essay purportedly written by a Black student, attitudes proved influential but motivation to control prejudice did not. Olson and Fazio (2007a) reported a similar pattern of findings regarding White participants' ratings of a Black relative to a White applicant to a volunteer position. Estimates of automatically activated racial attitudes derived from our standard priming procedure predicted the evaluations, but this relation was not moderated by either factor of the MCPR.

Why, with such a controllable judgment and no time pressure, did motivation essentially "fail" to act? We suspect that in order for motivation to control prejudice to influence one's race-related judgments, those judgments must be construed as race related. In Jackson's research, participants had the essay before them and their task was to evaluate it. In Olson and Fazio's research (2007a), a wealth of information about the targets was provided, including transcripts, a personal statement, and an interviewer's report. In both cases, it is likely that participants simply failed to consider the possibility that their judgments might be at all influenced by the race of the target person, and hence, no motivation was evoked. According to most models of social judgment, one of the first steps in correcting bias is noticing it (e.g., Wegener & Petty, 1995; Wilson & Brekke, 1994). The richness of the information that is available as a basis for one's judgment may render a potential source of bias relatively subtle and, hence, difficult to discover. Thus, the informational context can prevent motivation from being evoked.

Perspectives of the MODE Model on Contemporary Research

Early tests of the MODE model were conducted with the historic issue of attitude-behavior processes in mind, before implicit measures of attitudes and debates about their relation to explicit measures, conscious versus unconscious attitudes, and related issues examined in this volume came to the fore. As we have seen, the model has fared well in more recent tests employing an implicit measure of attitudes; the research has yielded supportive findings across a variety of judgmental and behavioral contexts. Given this success, it is our belief that the MODE model provides a valuable perspective concerning issues that have arisen regarding the meaning and interpretation of implicit measures. Below, we address some of these contemporary issues and consider how they are informed by the MODE model and our research findings.

Explicit Measures and Their Relation to Implicit Measures

A few years ago we reviewed what we referred to as the "burgeoning" research on implicit and explicit measures of attitudes (Fazio & Olson, 2003). As exemplified by this volume, this research has expanded to include a number of related research questions involving issues of

both measurement and the underlying cognitive processes presumably tapped by different measures. Myriad claims have been made about what these implicit measures actually measure (for extended reviews, see Fazio & Olson; Olson, 2003). As the popularity of implicit measurement tools has driven this research forward at a staggering pace, it is no irony that one of the age-old tools of attitude research—the questionnaire—has been thrust into the spotlight. As supposed measures of attitudes, it was natural for researchers to wonder whether implicit measures would reflect the same attitude reported on an explicit measure (for reviews, see Blair, 2001; Dovidio, Kawakami, & Beach, 2001). Thus, one question to come from implicit measurement research was this: What is the relation between implicit and explicit measures? Are they getting at the same thing?

The MODE model has a ready answer. Recall our earlier point that responses to a questionnaire necessarily involve overt expressions of one's attitudes. These responses are verbal behaviors. As such, they fall directly under the intended purview of the MODE model. Like any other behavior, verbal expressions can be affected by both one's automatically activated attitude toward the object and, opportunity willing, motivational factors. In the absence of motivation and opportunity, the evaluation automatically activated by the object should guide verbal responses. Thus, it is under conditions of low motivation or low opportunity (or both) that one should find correspondence between implicit and explicit measures, and this is precisely what a substantial body of research has shown.

The most obvious place to look for the MODE model's predictions of implicit-explicit correspondence is in research domains where motivational forces are likely to be completely absent. In early work (e.g., Fazio et al., 1986), the priming measure was employed to assess the automatic activation of attitudes toward objects for which there were few barriers to honest attitude reporting (e.g., puppies and cockroaches). This work indicated that just as explicit measures revealed a preference for puppies and a distaste for cockroaches, so did the (implicit) priming measure. There is little demand, either personally or socially, to adjust verbal reports of one's liking for such objects away from one's true attitudes. In contrast, research that has directly compared such attitude objects to more socially sensitive issues (e.g., abortion, Blacks, contraceptives, homosexuality, pornography) has observed far less correspondence between the implicit and explicit measures in the latter case (Dovidio & Fazio, 1992). Hence, we find that individuals' responses on explicit measures flow directly from the attitude the

object automatically evokes onto the questionnaire page when there is little motivation to report otherwise, but automatically activated attitudes are less likely to be openly expressed when the issues are more sensitive in nature. The same conclusion has been reached in research employing the IAT as an implicit measure of attitudes. For example, correspondence with explicit measures has been observed with respect to food and beverage preferences among other socially tame domains (Maison, Greenwald, & Bruin, 2001). More recent research, as well as meta-analyses, support the view that motivational factors presumably derived from social sensibilities moderate the relationship between the IAT and explicit measures (Hoffman, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, 2005).

Even in cases where one may be motivated to respond a certain way on an explicit measure, a lack of opportunity to do so will inhibit the influence of motivational processes, leaving only one's automatically activated attitudes to influence responses on an explicit measure. For example, Koole, Dijksterhuis, and van Knippenberg (2001) demonstrated that implicit and explicit measures of self-esteem correspond better when responses on the explicit measure are made very quickly or while distracted (see also Ranganath, Smith, & Nosek, 2008). Again, and as the MODE model predicts, correspondence between implicit and explicit measures appears when motivational factors have minimal impact on responses to the explicit measure.

According to the MODE model, such correspondence should be minimal or altogether absent to the extent that motivation and opportunity factors are both present. Again, this is what the research indicates. For example, in the motivationally volatile domain of racial prejudice in particular, where much of this research has been conducted, little correspondence has been observed between implicit and explicit measures (e.g., Fazio et al., 1995; Greenwald, McGhee, & Schwartz, 1998; Dovidio et al., 2002), and where it has been observed, correlations tend to be weak (e.g., McConnell & Liebold, 2001; Lepore & Brown, 1997; Wittenbrink, Judd, & Park, 1997). Of course, people vary in their motivation to control prejudice, and the MODE model would expect that dissociations between the two measure types would be observed only among individuals with relatively strong motivation to control prejudiced reactions. As we described earlier, this is precisely what Fazio and colleagues reported in 1995 with respect to an implicit priming measure of racial attitudes and the explicit Modern Racism Scale, well before debates about the relationship between implicit and explicit measures reached its current prominence.

Such moderating effects of the MCPR—with low-motivation respondents showing correspondence between implicit and explicit measures—have now been demonstrated across a variety of implicit and explicit measures (e.g., Banse, Seise, & Zerbes, 2001; Gawronski, Geschke, & Banse, 2003; Payne, 2001; Payne, Cheng, Govorun, & Stewart, 2005). For example, Payne et al. found the relation between estimates of racial attitudes assessed through their affect misattribution procedure and self-reported evaluations of Blacks to be moderated by MCPR scores. Similarly, Payne (2001) found that the relation between scores on the Modern Racism Scale and the extent to which performance in a weapons identification task was affected by Black versus White primes to vary as a function of MCPR scores. Correspondence was observed among those with lower motivation scores, but those higher in motivation to control prejudiced reactions displayed an inverse relation, indicative of motivated overcorrection.

Further evidence of the mixed processes predicted by the MODE model has accumulated from multiple laboratories and has demonstrated the joint influence of implicitly measured attitudes and motivation across a variety of measures and domains. For example, Florack, Scarabis, and Bless (2001) examined the relation between German students' attitudes toward Turkish immigrants, as assessed via an IAT, and their judgments of a Turkish juvenile delinquent whose criminal behavior was described in a lengthy newspaper article. The relation was moderated by scores on the need for cognition scale (Cacioppo, Petty, Feinstein, & Jarvis, 1996). The implicit measure proved predictive of assessments of the Turkish offender among those lower in need for cognition, but not among those more motivated to process information extensively. Dasgupta and Rivera (2006) assessed the impact of automatically activated prejudice toward gays and lesbians (also via an IAT) toward a gay interviewer. Consistent with the MODE model, this relation was influenced by two more motivationally oriented moderating variables, the extent to which participants endorsed egalitarian beliefs regarding gender and gender identity, and the extent to which they reported a commitment to monitoring and controlling their potentially prejudicial behavior. It was among individuals who lacked such motivation that a relation was observed.

Intriguingly, recent research by Scinta and Gable (2007) has yielded conceptually parallel findings in a very different domain: that of romantic relationships. Estimates of participants' automatically activated attitudes toward their romantic partners (obtained via an IAT in one study and an evaluative priming procedure involving the subliminal presentation

of images of the partner in a second study) were employed as predictors of reported satisfaction with the relationship. As expected, those with more positive attitudes reported greater satisfaction. However, this was true only among participants who were not entrapped in the relationship. Among participants who had invested heavily in the relationship and foresaw poor alternatives to their current partner, the relation between automatically activated attitudes toward the partner and relationship satisfaction was reversed. Using Scinta and Gable's terminology, such individuals faced formidable "barriers to exit" the relationship and, hence, were motivated to view the relationship favorably. Those characterized by relatively negative automatically activated attitudes and such motivational pressure actually reported the greatest satisfaction with the relationship. Indeed, across a variety of self-reports concerning satisfaction with the relationship, the moderating effects of barriers to exit yielded regression lines of the form depicted in panel B of Figure 2.2. Once again, then, we see evidence of differential correspondence between implicit measures (automatically activated attitudes toward the partner, in this case) and explicit measures (reports of relationship satisfaction) as function of motivational forces (barriers to exit the relationship).

This is not to say that the MODE model's two moderating factors, motivation and opportunity, can account for the universe of observed dissociations between implicit and explicit measures. Dissociations can occur for any number of reasons. For example, reiterating Azjen and Fishbein's (1977) argument regarding the importance of measuring attitudes and behavior at equivalent levels of specificity, Hofmann and colleagues (2005) pointed out that dissociations will occur if implicit and explicit measures do not correspond conceptually—one would not expect correspondence between two measures if they are measuring attitudes toward two different objects. Similarly, one would expect correspondence to be relatively weaker if the two measures fail to encourage the same categorization of the attitude object. Most people, objects, and issues are multiply categorizable, and contextual factors can promote one categorization over another. For example, if one measure were to encourage categorization of stimulus persons by race, but the other not, dissociations are more likely to be observed (see Fazio & Dunton, 1997; Olson & Fazio, 2003, for discussions of categorization by race).

Implications for Awareness of the Attitude

Observed dissociations between implicit and explicit measures have prompted many claims about the sorts of attitudes assessed by implicit

measures. One of the more provocative strikes us as also one of the least tenable: that individuals lack introspective access to their automatically activated attitudes. As we first argued a few years ago (Fazio & Olson, 2003), dissociations can easily be explained without invoking unawareness, and the use of an implicit measure does not, in and of itself, guarantee that individuals are unaware of their attitudes. Implicit measures simply provide estimates of individuals' attitudes without our having to ask them directly for such information. The measures themselves are silent with respect to the question of whether individuals are or are not aware of the evaluations they show evidence of associating with the attitude object (see Gawronski, Hofmann, & Wilbur, 2006, for a similar analysis and review). Indeed, the position of the MODE model stands contrary to the tenet that implicit measures reflect unconscious attitudes. Instead, the model maintains that people tend to generally be aware of their attitudes and that it is motivational forces, not some consciousness-impervious shield, that prevents their verbal expression. The evidence, as we describe next, supports the MODE model's perspective.

Perhaps the most direct way of testing the hypothesis that countervailing motivation, and not a lack of awareness, prevents the expression of automatically activated attitudes is to remove the source of motivation. Under such conditions, the explicit measure should then reflect the implicit measure. In other words, if one is aware of a given attitude, then an effective exhortation to report it honestly should reveal it. This is precisely what Nier (2005) demonstrated. Participants in his study completed both an IAT designed to assess racial prejudice and the Modern Racism Scale. Some participants completed the MRS under bogus pipeline conditions, where they were led to believe that the experimenter could identify any dishonesties in reporting, thus effectively rendering motivation impotent to conceal respondents' underlying sentiments. When participants completed the MRS under standard circumstances, no implicit-explicit correspondence was observed. However, when participants had been induced to be honest, IAT and MRS measures of racial prejudice showed clear correspondence (r = .51). Such a pattern would not have been apparent had participants been unaware of their implicit racial biases.

Additional evidence that people are aware of their implicitly measured attitudes comes from research demonstrating motivated correction for those attitudes, which, of course, is precisely what the research we reviewed earlier shows. For example, motivated participants in Olson and Fazio's (2004) study on first impressions of Blacks and Whites actively corrected for their implicitly assessed prejudice

according to their initial level of prejudice; motivated individuals with a greater degree of prejudice showed greater correction. Such corrective measures could not be taken had participants been unaware of the existence of their biases. Indeed, theories of bias correction in social judgments explicitly highlight the essential role that awareness of one's biases plays in corrective processes (e.g., Wegener & Petty, 1995; Wilson & Brekke, 1994). It is upon the suspicion that their judgments are being unduly influenced by some biasing force that individuals engage in correction. That such corrective processes have been observed repeatedly in the MODE model research suggests that people are aware of their implicitly measured attitudes. Importantly, we do not argue that people are always correct in estimating their biases, and it is interesting that the consistent pattern of overcorrection that we and others have observed time and again suggests that some Whites might actually overestimate the magnitude of influence exerted by their automatically activated attitudes toward Blacks.

We also do not claim that attitudes cannot form unconsciously—indeed, we have provided replicable evidence that attitudes can form and change unconsciously (Olson & Fazio, 2001, 2002, 2006). Evaluative conditioning can lead to the development of automatically activated attitudes that reflect pairings to which individuals have been exposed, even if they are unable to report any awareness of those pairings. In one such experiment (Olson & Fazio, 2002), participants underwent our evaluative conditioning procedure, and immediately thereafter their attitudes toward the conditioned stimuli (CS) were assessed via a subliminal priming measure. As a result, they never had any reason to consider their attitudes toward the CS and were not even aware of the presence of the CS during the performance of the implicit measure. Nevertheless, they showed evidence of greater positivity having been activated in response to the positive CS than the negative CS. Thus, the entire process, from attitude formation to attitude activation, can occur outside of conscious awareness. Yet, as demonstrated by the finding that explicit measures of attitude toward the CS also are sensitive to the evaluative conditioning (Olson & Fazio, 2001), once attention is called to feelings about the objects, such attitudes can be reported.

Unconscious processes unquestionably play a role in judgment and behavior. In any given instance, people may be unaware that their attitudes have been activated from memory. Moreover, they may be unaware that their attitudes are exerting some influence on their construals of the object in the immediate situation, or unaware of the magnitude of that influence. And there certainly is no reason to believe that

individuals are necessarily cognizant of the origins of their attitudes, that is, why they came to like or dislike some object. Yet, none of these possibilities necessitates the inference that individuals lack awareness of their attitudes per se.

Implicit Measures of Self-Esteem: A Case in Point

The MODE model has analogous implications for implicitly measured self-esteem, another domain in which lofty claims have been made about implicit measures' ability to penetrate the unconscious. Similar to claims of "unconscious prejudice," researchers have argued that the self-esteem tapped by implicit measures is unconscious and independent of one's conscious self-views, an argument bolstered by the low correlations observed between implicit and explicit measures of self-esteem (e.g., Bosson, Swann, & Pennebaker, 2000; Kernis, 2003; Koole and Pelham, 2003). Do we all have two independent attitudes toward the self, one conscious and one unconscious?

Again, the MODE model's perspective on the separation between implicit and explicit measures of self-esteem has little to do with unconsciousness versus consciousness or the existence of two independent self-attitudes. Instead, we argue that one's automatic self-views are accessible to conscious awareness and that motivated processes often interfere with forthright reporting on explicit measures. That is, explicit measures of self-esteem reflect not only automatically activated self-evaluations but also downstream forces. Our view is similar to that put forth by Dijksterhuis, Albers, and Bongers (see Chapter 8, this volume), who described implicit measures as more likely to reach "core" self-esteem, with responses on explicit measures more likely to be colored by a variety of motives.

One particularly strong motive is to view oneself in a positive light. Western culture in particular aims to champion and empower the individual, resulting in the social expectation that one view the self positively. Indeed, the mean response to traditional explicit measures of self-esteem is typically near the ceiling of the scale. On the other hand, modesty is another valued trait, and it is likely that some individuals strategically underpresent themselves in service of this motive, particularly in non-Western cultures (Kitayama & Uchida, 2003). The former motive might cause one to appear more positive toward the self than one's automatic reactions toward the self might imply. The latter, on the other hand, would encourage explicit reports that are more negative than those assessed implicitly.

Across two studies, we demonstrated that motivational forces such as these, and not separate conscious and unconscious self-attitudes, underlie dissociations between implicit and explicit measures of self-esteem (Olson, Fazio, & Hermann, 2007). First, we asked respondents to rate themselves on a variety of trait variables relating to over- (e.g., proud, boastful) and underpresenting (e.g., modest, meek) styles after completing implicit (IAT) and explicit (Rosenberg, feeling thermometer) measures of self-esteem. As expected, those who reported more positive self-views on the explicit measure than their implicitly measured self-esteem would imply admitted to being more proud and boastful, and less modest and meek (see Lambird & Mann, 2006, for a similar analysis). In a second study, we directly manipulated the operation of participants' self-presentational motives. Some participants completed implicit and explicit measures under the usual conditions, without any special instructions. Others were implored to be honest when completing the explicit measure. More specifically, they were told that when answering self-related questions, some people tend to overpresent themselves by being proud or boastful and others tend to underpresent themselves by being modest or humble. They were urged to do neither. In the control condition, little correspondence was found between the two measure types, replicating much past research. However, when urged to be honest, participants' implicit and explicit responses showed greater convergence. It would be difficult to explain this pattern of relations if one were to assume the existence of two independent self-attitudes, one of which was hidden from introspective access.

This is not to say that dissociations between implicit and explicit measures of self-esteem are not interesting or important. Quite the contrary, the presence of a discrepancy can serve as a very informative marker. Pursuing such an approach, Jordan and colleagues (e.g., Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003; see Chapter 9, this volume) have provided a fascinating window into individuals characterized by relatively high scores on an explicit measure of self-esteem, but relatively low scores on an implicit measure. These individuals exhibit what can be termed a narcissistic, fragile, or defensive form of high self-esteem (see also Kernis, Abend, Goldman, Shrira, Paradise, & Hampton, 2005). Their defensiveness can lead to outgroup derogation, pronounced dissonance reduction effects, and more (see also Jordan, Spencer, & Zanna, 2005). We would argue that the high self-esteem such individuals report on the explicit measure of self-esteem is a product of this very same defensive style. Thus, it is not the discrepancy per se that motivates their defensive reactions. Instead, the discrepancy is itself symptomatic of the defensive style with which they respond to their less-than-desired automatically activated self-views and to self-related threats more generally. Our point is that, as verbal behaviors, responses to explicit measures of self-esteem are necessarily an exercise in self-presentation. Hence, they may not offer an accurate portrait of automatically activated self-evaluations, but instead may be behavioral manifestations of additional phenomena, such as narcissism. From the perspective of the MODE model, it is the schism between the attitude that is activated automatically and more motivated processes that produce the behavior under study, not some schism between the conscious and the unconscious.

Single versus Dual Attitudes

As should be abundantly obvious by now, the MODE model stands in contrast to the stance of models that postulate the existence of dual representations of attitude in memory (e.g., Wilson, Lindsey, & Schooler, 2000). Our perspective does not view responses to an explicit measure as indicative of a representation in memory that is distinct from the automatically activated attitude. Instead, such responses are viewed as verbal behaviors that, under appropriate circumstances, can be influenced by considerations other than, or in addition to, the automatically activated attitude. We continue to believe, just as we argued earlier (Fazio & Olson, 2003), that "implicit" and "explicit" are best viewed as properties of the measure, not properties of the construct that is being measured. Such terms as "implicit attitude" and "explicit attitude," and "implicit self-esteem" and "explicit self-esteem," invite misinterpretation, because they imply the existence of dual representations. Within the context of the MODE model, such references are inappropriate; the model views automatically activated attitudes as the representations captured by an implicit measure and as the starting point for verbal responses to an explicit measure.

Implications for Predicting Behavior

A consensus appears to be emerging from recent research on implicit and explicit attitudinal processes. Most succinctly, it is the view that implicit attitudes predict automatic behavior, and explicit attitudes predict controlled behavior. Findings by Dovidio and colleagues (1997, 2002), McConnell and Liebold (2001), Neumann, Hülsenbeck, and Seibt (2004), and others seem to support such a view. Here and elsewhere,

implicit measures were found to correlate with nonverbal behavior like looking and smiling, whereas explicit measures related more strongly to controllable behaviors.

However, a look through the literature indicates that a mere "mapping" of implicit measures to uncontrollable behavior and explicit measures to controllable behavior is a simplistic characterization of the research findings. Clearly, implicit measures have been shown to predict more than the merely less-controllable classes of behavior. As we reviewed earlier, the priming measure of racial prejudice we employ has related to explicit judgments of the quality of a Black student's writing, reported first impressions of Black and White social targets, evaluations of job candidates, and the longevity of interracial roommate relationships. Other implicit measures employed in other labs have similarly shown that even behaviors fully susceptible to conscious control can be guided by the automatic processes assessed by implicit measures (e.g., Vanman, Saltz, Nathan, & Warren, 2004). Thus, the view that implicit measures (and the constructs they measure) are limited to the prediction of less-controllable behavior grossly underestimates their predictive power.

From the MODE model's perspective, the attitude that is automatically evoked upon encountering an object can, if unimpeded by motivation, determine behavior of all sorts, from approach/avoidance behavior to smiles and frowns to verbal declarations of admiration or disgust. For example, recall the research we reviewed earlier indicating that the influence of automatically activated attitudes begins early in the attention and perceptual process, coloring construals of the object and the situation at hand. These early influences can have a cascading effect on later judgments and behaviors, whether they be controlled or automatic in nature.

Such downstream consequences of automatic processes are particularly well illustrated by a class of research demonstrating stronger relations between automatically activated attitudes (as measured by some implicit measure) and some judgment or behavior when the individual's resources have been depleted in some way. For example, Hofmann, Rauch, and Gawronski (2007) assessed participants' automatic attitudes toward candy (using an IAT) as well as their personal "dietary restraint standards," that is, the extent to which they were motivated to monitor and control their diet. All participants were exposed to a dramatic movie scene, but those in the resource-depletion condition were instructed to suppress the expression of emotion while viewing the clip. Later, in a product testing phase of the experiment, participants were

provided with an opportunity to consume candy. Just as the MODE model would predict, resource-depleted participants' candy consumption was primarily predicted by their automatic candy attitudes, whereas those with fuller resources ate according to their dietary standards. Similar findings supportive of the MODE model are now beginning to appear in the substance use and abuse literature (see Wiers & Stacy, 2006, for an overview). For example, Hofmann, Gschwendner, Friese, and Schmitt (2007) reported that automatically activated attitudes toward alcohol predict consumption behavior particularly when cognitive resources are diminished from an ego-demanding task. Govorun and Payne (2006) demonstrated that a resource-depleting Stroop color-naming task diminished the controlled component of a process-dissociation procedure in a weapons identification task, leaving a greater role for automatic processes to guide judgments. Essentially, these depletion manipulations represent manipulation of the opportunity factor. So, when the opportunity for downstream motivational forces to be influential is minimized, stronger relations between automatically activated attitudes and behavior emerge.

Conclusion

It was nearly two decades ago that the MODE model was first formalized (Fazio, 1990). We believe the theory continues not only to be relevant, but also to provide a general theoretical framework for addressing some of the more contemporary questions addressed in this volume. The model considers the evaluation automatically evoked upon encountering an object as the starting point of the perceptual, cognitive, and motivational processes that guide object-related behavior. It considers responses to explicit measures of attitudes a form of verbal behavior, comparable to other social judgments. As such, they are susceptible to the same sorts of influences of motivated processes that, opportunity permitting, can steer one's behavior away from that implied by the attitude.

The MODE model provides input on many of the issues discussed in this volume, including the relation between implicit and explicit measures, the nature of the constructs tapped by each, the role of the unconscious in attitudes, and the sorts of important social behaviors guided by the various attitudinal processes. Our view is that many of these debates are not new; they have parallels in earlier literature regarding the attitude-behavior relation and, hence, can be informed

by earlier theory and research. The advent of implicit measures has added important tools by which these questions can be pursued with greater precision and rigor. However, it takes theory to ground the rapidly accumulating research findings to basic principles, which is what we believe the MODE model provides.

References

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888–918.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice Hall.

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173–222). Mahwah, NJ: Erlbaum.

Ajzen, I., & Sexton, J. (1999). Depth of processing, belief congruence, and attitude-behavior correspondence. In S. Chaiken & Y. Trope (Ed.), *Dual-process theories in social psychology* (pp. 117–138). New York: Guilford Press.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift fur Experimentelle Psychologie, 48,* 145–160.

Baumeister, R. F., Bratlavasky, M., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality & Social Psychology, 74,* 1252–1265.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117,* 497–529.

Biernat, M., & Vescio, T. K. (1993). Categorization and stereotyping: Effects of group context on memory and social judgment. *Journal of Experimental Social Psychology, 29,* 166–202.

Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the legacy and future of social cognition* (pp. 359–374). Mahwah, NJ: Erlbaum.

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality & Social Psychology, 79,* 631–643.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64,* 123–152.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, B. W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in the need for cognition. *Psychological Bulletin, 119,* 197–253.

Carretta, T. R., & Moreland, R. L. (1982). Nixon and Watergate: A field demonstration of belief perseverance. *Personality & Social Psychology Bulletin, 8,* 446–453.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York: Guilford Press.

Dasgupta, N., & Rivera, L. M. (2006). From automatic antigay prejudice to behavior: The moderating role of conscious beliefs about gender and behavioral control. *Journal of Personality & Social Psychology, 91,* 268–280.

DePaulo, B. M., & Friedman, H. S. (1998). Nonverbal communication. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 3–40). New York: Random House.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality & Social Psychology, 56,* 5–18.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 204–237). New York: Russell Sage Foundation.

Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 3–32). Newbury Park, CA: Sage.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Vol. 4. Intergroup processes* (pp. 175–197). Malden, MA: Blackwell.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interactions. *Journal of Personality & Social Psychology, 82,* 62–68.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510–540.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality & Social Psychology Bulletin, 23,* 316–326.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Fort Worth, TX: Harcourt Brace Jovanovich.

Fabrigar, L., Petty, R. E., Smith, S. M., & Crites, S. L. (2006). Understanding knowledge effects on attitude-behavior consistency: The role of relevance, complexity, and amount of knowledge. *Journal of Personality and Social Psychology, 90,* 556–577.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego: Academic Press.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion, 15,* 115–141.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition, 25,* 664–703.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 33,* 451–470.

Fazio R. H., & Hilden, L. (2001). Emotional reactions to a seemingly prejudiced response. *Personality & Social Psychology Bulletin, 27,* 538–549.

Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology, 51,* 505–514.

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

Fazio, R. H., Ledbetter, J. E., & Towles-Schwen, T. (2000). On the costs of accessible attitudes: Detecting that the attitude object has changed. *Journal of Personality and Social Psychology, 78,* 197–210.

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

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

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 97–116). New York: Guilford Press.

Florack, A., Scarabis, M., & Bless, H. (2001). When do associations matter? The use of implicit associations toward ethnic groups in person judgments. *Journal of Experimental Social Psychology, 37,* 518–524.

Gaertner, S. L., & Dovidio, J. F. (1977). The subtlety of white racism, arousal and helping behavior. *Journal of Personality & Social Psychology, 35,* 691–707.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology, 33,* 573–589.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition, 15,* 485–499.

Govorun, O., & Payne, B. K. (2006). Ego depletion and prejudice: Separating automatic and controlled components. *Social Cognition, 24,* 111–136.

Greenwald, A. G., McGhee, D., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Task. *Journal of Personality & Social Psychology, 74,* 1469–1480.

Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49,* 129–134.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94,* 319–340.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles.* New York: Guilford Press.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality & Social Psychology Bulletin, 31,* 1369–1385.

Hofmann, W., Gschwendner, T., Friese, M., & Schmitt, M. (2007, January). *Impulsive and reflective paths to behavior: Available control resources moderate the predictive validity of implicit and explicit attitudes.* Paper presented at the annual meeting of the Society of Personality and Social Psychology, Memphis, TN.

Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology, 43,* 497–504.

Houston, D. A., & Fazio, R. H. (1989). Biased processing as a function of attitude accessibility: Making objective judgments subjectively. *Social Cognition, 7,* 51–66.

Jackson, J. R. (1997). *Automatically activated racial attitudes.* Unpublished doctoral dissertation, Indiana University.

Jamieson, D. W., & Zanna, M. P. (1989). Need for structure in attitude formation and expression. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 383–406). Hillsdale, NJ: Erlbaum.

Jordan, C. H., Spencer, S. J., & Zanna, M. P. (2005). Types of high self-esteem and prejudice: How implicit self-esteem relates to racial discrimination among high explicit self-esteem individuals. *Personality & Social Psychology Bulletin, 31,* 693–702.

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive self-esteem. *Journal of Personality & Social Psychology, 85,* 969–978.

Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality & Social Psychology, 52,* 536–546.

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality & Social Psychology, 55,* 893–905.

Kernis, M. H. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry, 14,* 1–26.

Kernis, M. H., Abend, T. A., Goldman, B. M., Shrira, I., Paradise, A. N., & Hampton, C. (2005). Self-serving responses arising from discrepancies between explicit and implicit self-esteem. *Self and Identity, 4,* 311–330.

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

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality & Social Psychology, 80,* 669–685.

Koole, S. L., & Pelham, B. W. (2003). On the nature of implicit self-esteem: The case of the name-letter effect. In S. Spencer, S. Fein, & M. P. Zanna (Eds.), *Motivated social perception: The Ontario Symposium* (pp. 93–116). Hillsdale, NJ: Lawrence Erlbaum.

Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases.* New York: Plenum Press.

Lambird, K. H., & Mann, T. (2006). When do ego threats lead to self-regulation failure? Negative consequences of defensive high self-esteem. *Personality & Social Psychology Bulletin, 32,* 1177–1187.

Lepore L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality & Social Psychology, 72,* 275–87.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality & Social Psychology, 37,* 2098–2109.

Maison, D., Greenwald, A.G., & Bruin, R. (2001). The Implicit Association Test as a measure of consumer attitudes. *Polish Psychological Bulletin, 2,* 61–79.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic Press.

McConnell, A. R., & Liebold, J. M. (2001). Relations between the Implicit Association Test, explicit racial attitudes, and discriminatory behavior. *Journal of Experimental Social Psychology, 37,* 435–442.

Neumann, R., Hülsenbeck, K., & Seibt, B. (2004). Attitudes towards people with AIDS and avoidance behavior: Automatic and reflective bases of behavior. *Journal of Experimental Social Psychology, 40,* 543–550.

Nier, J. A. (2005). How dissociated are implicit and explicit racial attitudes? A bogus pipeline approach. *Group Processes & Intergroup Relations, 8,* 39–52.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565–584.

Olson, M. A. (in press). Measures of prejudice. In T. Nelson (Ed.), *Handbook of Prejudice.* Hillsdale, NJ: Erlbaum.

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

Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition, 20,* 89–103.

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

Olson, M. A., & Fazio, R. H. (2004). Trait inferences as a function of automatically activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology, 26,* 1–12.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality & Social Psychology Bulletin, 32,* 421–433.

Olson, M. A., & Fazio, R. H. (2007a). Discordant evaluations of Blacks affect nonverbal behavior. *Personality & Social Psychology Bulletin,* 33, 1214–1224.

Olson, M. A., & Fazio, R. H. (2007b). Unpublished raw data. University of Tennessee.

Olson, M. A., Fazio, R. H., & Hermann, A. D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science, 18,* 287–291.

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

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

Petty, R. E., & Krosnick, J. A. (Eds.). (1995). *Attitude strength: Antecedents and consequences.* Hillsdale, NJ: Erlbaum.

Petty, R. E., & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 41–72). New York: Guilford Press.

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology, 44,* 386–396.

Roskos-Ewoldsen, D. R., & Fazio, R. H. (1992). On the orienting value of attitudes: Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality & Social Psychology, 63,* 198–211.

Sanbonmatsu, D. M., & Fazio, R. H. (1990). The role of attitudes in memory-based decision making. *Journal of Personality & Social Psychology, 59,* 614–622.

Schuette, R. A., & Fazio, R. H. (1995). Attitude accessibility and motivation as determinants of biased processing: A test of the MODE model. *Personality & Social Psychology Bulletin, 21,* 704–710.

Scinta, A., & Gable, L. (2007). Automatic and self-reported attitudes in romantic relationships. *Personality and Social Psychology Bulletin, 33,* 1008–1022.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology, 29,* 209–269.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality & Social Psychology, 71,* 888–898.

Snyder, M., & Gangestad, S. (1982). Choosing social situations: Two investigations of self-monitoring processes. *Journal of Personality & Social Psychology, 43,* 123–135.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220–247.

Towles-Schwen, T., & Fazio, R. H. (2001). On the origins of racial attitudes: Correlates of childhood experiences. *Personality & Social Psychology Bulletin, 27,* 162–175.

Towles-Schwen, T., & Fazio, R. H. (2003). Choosing social situations: The relation between automatically activated racial attitudes and anticipated comfort interacting with African Americans. *Personality & Social Psychology Bulletin, 29,* 170–182.

Towles-Schwen, T., & Fazio, R. H. (2006). Automatically activated racial attitudes as predictors of the success of interracial roommate relationships. *Journal of Experimental Social Psychology, 42,* 698–705.

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

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality & Social Psychology, 68,* 36–51.

Wiers, R. W & Stacy, A. W. (Eds.). (2006). *Handbook of implicit cognition and addiction.* Thousand Oaks, CA: Sage Publishers.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

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. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality & Social Psychology, 72,* 262–274.