

In B. Gawronski, & F. Strack (Eds.) (2012). Cognitive consistency: A fundamental principle in social cognition. New York: Guilford.

18

## Cognitive Consistency in Prejudice-Related Belief Systems

Integrating Old-Fashioned, Modern,  
Aversive, and Implicit Forms of Prejudice

Bertram Gawronski  
Paula M. Brochu  
Rajees Sritharan  
Fritz Strack

A central theme in current research on prejudice is the increasingly subtle nature of negative responses to members of stigmatized groups. For instance, with regard to racial prejudice in the United States, it has been argued that prejudice against African Americans has simply changed its face instead of disappearing. This conclusion is based on the finding that, even though overt expression of prejudiced beliefs about racial groups has steadily declined over the past decades (e.g., Schuhman, Steeh, Bobo, & Kyrstan, 1997; Sniderman & Carmines, 1997), levels of racial conflict between groups have changed only to a small extent (Zárate, 2009).

To account for these discrepancies, social psychologists have argued that the increasing acceptance of egalitarian values made the overt endorsement

of prejudiced beliefs socially unacceptable. Yet negative sentiments against minority groups are assumed to linger in more subtle forms that differ from traditional, blatant forms of prejudice. The most influential constructs in this regard are *modern prejudice* (McConahay, 1986), *aversive prejudice* (Gaertner & Dovidio, 1986), and *implicit prejudice* (Rudman, Greenwald, Mellott, & Schwartz, 1999).

Despite shared origin in the assumption that prejudice continues to prevail in subtle forms, there have been hardly any attempts to clarify the conceptual relations between the proposed constructs (for some notable exceptions, see Nail, Harton, & Decker, 2003; Son Hing, Chung-Yan, Hamilton, & Zanna, 2008). In this chapter, we review a theoretical model that uses the basic principles of cognitive consistency to integrate old-fashioned, modern, aversive, and implicit prejudice in a single framework (Gawronski, Peters, Brochu, & Strack, 2008; see also Brochu, Gawronski, & Esses, 2008). Adopting a consistency perspective not only clarifies the relations between blatant and various subtle forms of prejudice; it also provides deeper insights into the dynamic processes that can make prejudice reduction rather difficult. Toward this end, we first review the concepts of old-fashioned, modern, aversive, and implicit prejudice. We then present our integrative prejudice framework, discussing how a cognitive consistency perspective can advance research on prejudice through (1) the integration of different forms of prejudice in a single framework, (2) the derivation of novel predictions about the relation between implicit and explicit prejudice, (3) the discovery of a previously undetected type of prejudice, and (4) the identification of potential obstacles in prejudice reduction.

### DIFFERENT FORMS OF RACIAL PREJUDICE

The distinction between blatant and subtle forms of prejudice has its roots in a puzzling discrepancy between public opinion polls and racial conflicts in North America. Whereas public opinion polls have suggested a steady decline in negative beliefs about racial minority groups since World War II, racial conflicts have shown only a moderate reduction (Greeley & Sheatsley, 1971; Taylor, Sheatsley, & Greeley, 1978). Several theorists explain this dissociation by the emergent conflict between *old-fashioned racism* and the increasing importance of egalitarian values (Gaertner & Dovidio, 1986; Katz & Hass, 1988; McConahay, 1986; Sears, 1988). According to these theorists, old-fashioned racism is characterized by the overt endorsement of nonegalitarian beliefs, such as support for racial segregation and open discrimination. These nonegalitarian beliefs stand in contrast to the ideal of equality and equal opportunity, which led to a decline in old-fashioned prejudice with the increasing acceptance of egalitarian values. However, this decline did not necessarily lead to a corresponding reduction in the negative senti-

ments against racial minorities that still seem to linger under the surface of egalitarianism (Zárate, 2009).

According to theories of *modern racism*, the conflict between egalitarian values and negative sentiments simply led to a change in the “channels” through which racial prejudice is expressed. Instead of supporting racial segregation and open discrimination, modern racists are claimed to express their negative sentiments against racial minorities in the belief that racial discrimination no longer exists. More specifically, modern racism is proposed to be rooted in four interrelated belief-components (McConahay, 1986, p. 92):

- (1) Discrimination is a thing of the past because Blacks now have the freedom to compete in the marketplace and to enjoy those things they can afford.
- (2) Blacks are pushing too hard, too fast and into places where they are not wanted.
- (3) These tactics and demands are unfair.
- (4) Therefore, recent gains are undeserved and the prestige granting institutions of society are giving Blacks more attention and the concomitant status than they deserve.

The critical difference between modern and old-fashioned racism is that people who hold these discriminatory beliefs do not regard them as expressions of racial prejudice. Instead, modern racists tend to interpret these beliefs as unbiased reflections of empirical facts that are unrelated to personal sentiments about racial groups.

Whereas theories of modern racism address the expression of racial prejudice through discriminatory beliefs that are deemed nonprejudicial by the individual, the concept of *aversive racism* involves internal conflicts resulting from genuine internalizations of egalitarian values (Dovidio & Gaertner, 2004; Gaertner & Dovidio, 1986). According to aversive racism theory, people often experience a conflict between negative feelings toward racial minority members and a personal desire to be nonprejudiced. This conflict is assumed to result in a state of attitudinal ambivalence, which is described as “a particular type of ambivalence in which the conflict is between feelings and beliefs associated with a sincerely egalitarian value system and unacknowledged negative feelings and beliefs about blacks” (Gaertner & Dovidio, 1986, p. 62). An important aspect of aversive racism is that these feelings are not necessarily hostile or vicious. Instead, the negative feelings held by aversive racists are assumed to involve discomfort, uneasiness, or fear, which may lead to avoidance behavior despite the personal importance of egalitarian values.

With the development of indirect measurement procedures—such as the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) and evaluative priming (Fazio, Jackson, Dunton, & Williams, 1995)—research on subtle forms of prejudice became heavily influenced by the concept of *implicit prejudice*.<sup>1</sup> A common assumption in this research is that indirect mea-

surement procedures are capable of assessing mental contents that people are either unwilling (Fazio et al., 1995) or unable (Greenwald & Banaji, 1995) to report on direct self-report measures (but see Gawronski, LeBel, & Peters, 2007, for a critical evaluation of this assumption). In line with this contention, many researchers distinguish between implicit prejudice (as assessed by indirect measures) and explicit prejudice (as assessed by direct measures), with the former being interpreted as a subtle variant of prejudice that persists despite the increasing impact of egalitarian values. Even though research on implicit prejudice rarely distinguishes between different kinds of explicit prejudice, the rather common dissociations between direct and indirect measures have been integrated into the theory of aversive racism (Dovidio & Gaertner, 2004). Specifically, it is assumed that indirect measures reveal the negative feelings experienced by aversive racists, which may be in conflict with explicitly endorsed egalitarian values that influence responses on direct self-report measures (e.g., Son Hing, Chung-Yan, Grunfeld, Robichaud, & Zanna, 2005).

### THE INTEGRATIVE PREJUDICE FRAMEWORK

Research and theorizing on old-fashioned, modern, aversive, and implicit prejudice have been somewhat selective in the sense that there have been no attempts to provide a comprehensive framework that integrates all of the proposed constructs. To overcome this limitation, Gawronski, Peters, et al. (2008) proposed an integrative model that uses the basic principles of cognitive consistency to specify the relations between old-fashioned, modern, aversive, and implicit prejudice. In addition to explaining how the four kinds of prejudice are related to each other, their integrative prejudice framework implies novel predictions about the relation between implicit and explicit prejudice, which have been empirically confirmed in several independent studies. The functional value of the model is further supported by the discovery of a previously undetected form of prejudice, and by its capacity to clarify various obstacles in prejudice reduction. In the following sections, we first explain the theoretical basis of the model and then discuss each of these issues in turn.

#### Associative versus Propositional Processes

The conceptual core of our integrative prejudice framework is the distinction between *affective reactions* and *evaluative judgments* as two qualitatively distinct kinds of evaluative response. Drawing on generalized dual-process models (Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004), we argue that the two kinds of evaluative response have their roots in two distinct, yet interacting, mental processes: associative and propositional processes. Specifically, we assume that people tend to experience a positive or nega-

tive affective reaction to a given object depending on the evaluative quality of the associations that are activated in response to that object. The affective reactions resulting from activated associations usually provide the basis for a verbally endorsed evaluative judgment (e.g., a negative affective reaction toward object X being translated into the propositional evaluation "I dislike X"), unless the evaluation implied by the affective response is inconsistent with other relevant information. More generally, we argue that the perceived validity of a cognitive element—such as the evaluation implied by an affective reaction—depends on the consistency of this element with other elements considered to be relevant for a given judgment. In the case of evaluative judgments, such elements may include nonevaluative beliefs referring to states of affairs in the world or evaluative beliefs about other attitude objects (Jones & Gerard, 1967). If the evaluation implied by an affective reaction is consistent with all other relevant beliefs, it is typically considered valid and may thus serve as the basis for an evaluative judgment. If, however, the evaluation implied by an affective reaction is inconsistent with other relevant beliefs, the inconsistency will have to be resolved in order to avoid aversive feelings of cognitive dissonance (Festinger, 1957). One possible way to achieve consistency is to reject—or invalidate—the evaluation implied by the affective reaction as a basis for an evaluative judgment. However, merely rejecting an affective reaction as a basis for an evaluative judgment does not necessarily deactivate the associations that gave rise to this reaction (Deutsch, Gawronski, & Strack, 2006). Hence, inconsistency-related rejections of affective reactions often lead to dissociations between affective reactions and evaluative judgments, such that affective reactions tend to persist despite inconsistency-related changes in evaluative judgments (e.g., Gawronski & Strack, 2004).

#### Consistency Principles in Prejudice-Related Belief Systems

The notion of consistency-based validity assessments serves as the theoretical basis for our integrative prejudice framework (Gawronski, Peters, et al., 2008). To illustrate how this idea explains the conceptual relation between old-fashioned, modern, aversive, and implicit prejudice, imagine a case in which an individual experiences negative affective reactions toward Black people. In terms of our model, these reactions can be understood as the product of the evaluative associations activated in response to Black people (e.g., associations linking Black people to negative social stereotypes). To the extent that these associations are the driving force underlying people's responses on indirect measures (e.g., Fazio et al., 1995; Greenwald et al., 1998), the affective responses resulting from activated associations can be equated with what many researchers describe as *implicit prejudice* (e.g., Dovidio & Gaertner, 2004).

According to our model, negative affective responses to Black people may provide the basis for a negative evaluative judgment (i.e., "I dislike

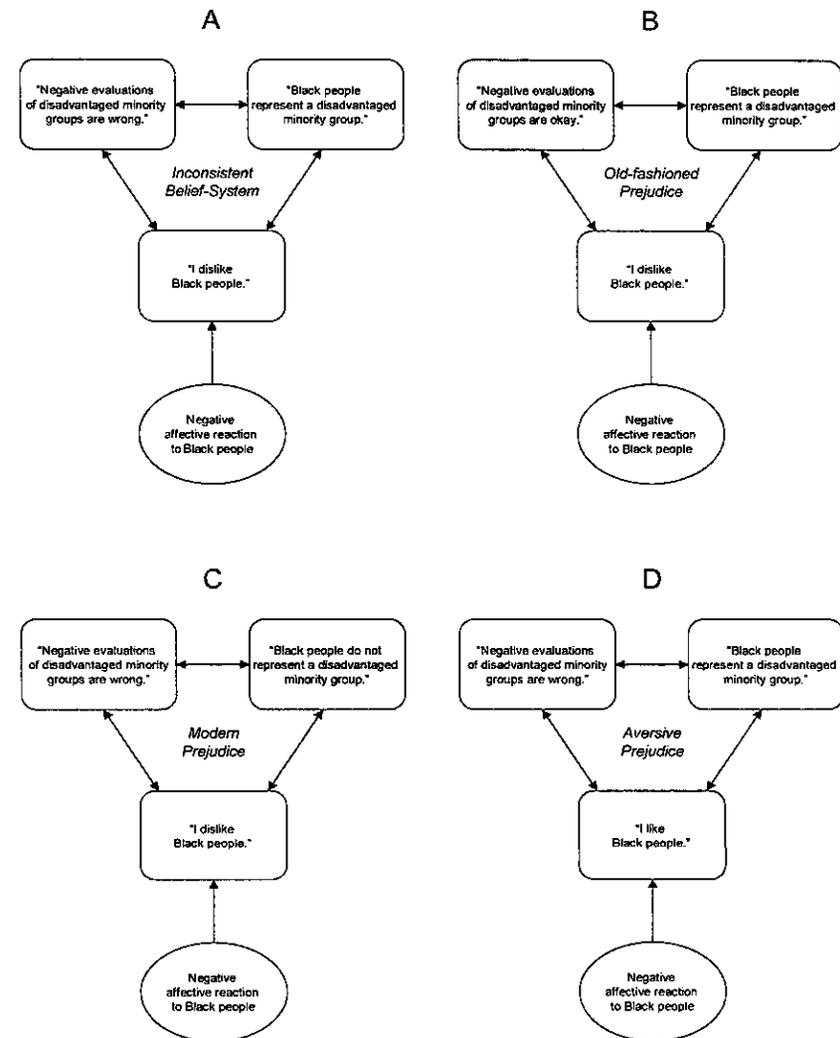
Black people"). However, whether or not the evaluation implied by the affective response will be explicitly endorsed depends on the consistency of this evaluation with other beliefs considered to be relevant. In the current example, such beliefs may include perceived discrimination against Black people and egalitarianism-related, nonprejudicial goals.<sup>2</sup> Specifically, the resulting set of judgment-relevant elements may include the following three propositions (see Figure 18.1, Panel A):

1. "I dislike Black people."
2. "Black people represent a disadvantaged minority group."
3. "Negative evaluations of disadvantaged minority groups are wrong."

These three propositions are inconsistent with each other in that they cannot be endorsed simultaneously without violating the basic notion of cognitive consistency (Festinger, 1957). Proposition 1 is inconsistent with the joint implication of Propositions 2 and 3; Proposition 2 is inconsistent with the joint implication of Propositions 1 and 3; and Proposition 3 is inconsistent with the joint implication of Propositions 1 and 2. Thus, to avoid aversive feelings of cognitive dissonance (Festinger, 1957), this set of beliefs has to be revised by either rejecting one of the three propositions as false or by searching for a new proposition that resolves the inconsistency (Gawronski & Strack, 2004). For the current analysis, the most important strategy is the rejection of one of the three propositions. The three possible cases implied by this strategy are illustrated in Figure 18.1 (Panels B, C, D).

First, people may reject the proposition representing egalitarianism-related, nonprejudicial goals (e.g., "Negative evaluations of disadvantaged minority groups are okay"). In this case, people may still agree that Black people represent a disadvantaged minority group. However, this belief does not result in a rejection of negative affective reactions as a basis for evaluative judgments, because negative judgments of disadvantaged minority groups are considered acceptable. Thus, negative affective reactions to Black people may directly translate into negative judgments (see Figure 18.1, Panel B). In our framework, this pattern reflects the notion of *old-fashioned prejudice*, which is characterized by the rejection of nonprejudicial, egalitarian values and open support for social discrimination (McConahay, 1986).

Second, people may reject the proposition representing discriminatory beliefs (e.g., "Black people do not represent a disadvantaged minority group"). In this case, people may still hold strong egalitarianism-related, nonprejudicial goals. However, these goals do not result in a rejection of negative affective reactions as a basis for evaluative judgments, because Black people are not considered to be targets of discrimination (Franco & Maass, 1999). Accordingly, negative affective reactions may directly translate into negative judgments (see Figure 18.1, Panel C). This pattern reflects the notion of *modern prejudice*, which is characterized by the denial that there is contin-



**FIGURE 18.1.** Interplay between affective reactions (circles) and propositional beliefs (squares) in racial prejudice against Black people. Panel A depicts the case of an inconsistent belief system; Panels B, C, and D depict consistent belief systems, reflecting different forms of racial prejudice. Adapted from Gawronski, Peters, Brochu, and Strack (2008). Copyright 2008 by Sage Publications, Inc. Adapted by permission.

ued social discrimination (McConahay, 1986). From the perspective of our model, denying the existence of racial discrimination can legitimize negative responses to Black people despite the endorsement of egalitarian values.

Third, people may reject their negative affective reactions as a valid basis for an evaluative judgment (e.g., "I like Black people"). Such a rejection may occur when people hold strong egalitarianism-related, nonprejudicial goals and, at the same time, agree that Black people represent a disadvantaged minority group. In this case, negative affective reactions to Black people will not translate into negative evaluative judgments (see Figure 18.1, Panel D). Rather, people's evaluative judgments will be relatively neutral (or even positive), irrespective of the valence of their affective reactions. As we outlined earlier, however, the mere rejection—or invalidation—of affective reactions as a basis for evaluative judgments does not necessarily deactivate the associations that gave rise to these reactions (Deutsch et al., 2006). Thus, even though negative affective reactions may not be reflected in evaluative judgments, spontaneous affective reactions may still be negative. In our framework, this pattern reflects the notion of *aversive prejudice*, such that people may experience negative feelings toward Black people even when they hold strong egalitarianism-related, nonprejudicial goals and, at the same time, believe that Black people represent a disadvantaged minority group (Dovidio & Gaertner, 2004).

### Empirical Evidence

Our integrative prejudice framework provides not only a conceptual integration of different forms of prejudice; it also implies novel predictions about the relation between implicit and explicit prejudice, with the former being conceptualized as negative affective reactions resulting from activated associations, and the latter as negative evaluative judgments reflected in direct self-reports. Specifically, our model predicts that the correlation between implicit and explicit prejudice should be moderated by the interaction of egalitarianism-related, nonprejudicial goals and perceptions of discrimination. In particular, people should base their evaluative judgments about a social group on their immediate affective reactions toward this group when they (1) believe that this group is a target of social discrimination but do not endorse egalitarianism-related, nonprejudicial goals (see Figure 18.1, Panel B) or (2) endorse egalitarianism-related, nonprejudicial goals but do not believe that this group is a target of social discrimination (see Figure 18.1, Panel C); in contrast, immediate affective reactions should *not* be reflected in people's evaluative judgments when they (3) hold strong egalitarianism-related, nonprejudicial goals and, at the same time, believe that the group is a target of social discrimination (see Figure 18.1, Panel D). In statistical terms, these predictions imply a significant positive correlation between implicit prejudice (as assessed by indirect measures; e.g., the Implicit Association Test or evaluative priming) and explicit prejudice (as assessed by direct measures; e.g., a

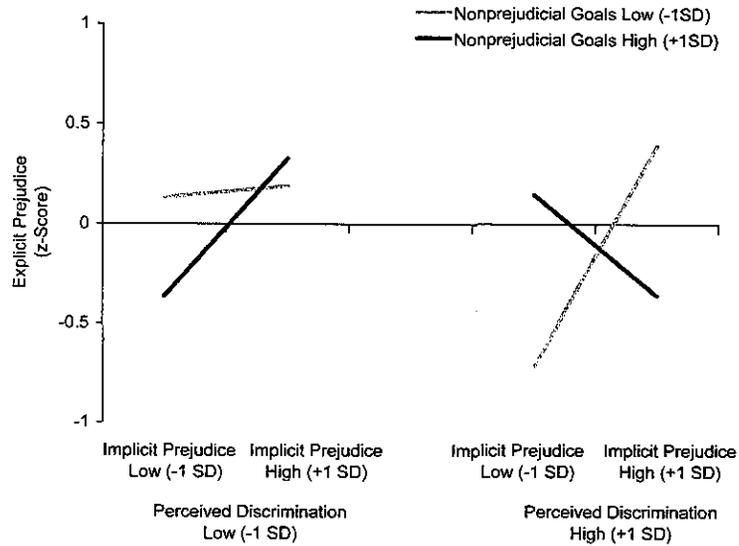
feeling thermometer or likability ratings) when (a) perceived discrimination is low and egalitarianism-related, nonprejudicial goals are strong, and (b) perceived discrimination is high and egalitarianism-related, nonprejudicial goals are weak; in contrast, implicit and explicit prejudice should be unrelated (or negatively related) when (c) perceived discrimination is high and, at the same time, egalitarianism-related, nonprejudicial goals are strong.

So far, these predictions have been confirmed in four independent studies using different target groups and multiple measures. Attesting to the domain-independence of our framework, the predicted pattern has been replicated for prejudice against Black people (Gawronski, Peters, et al., 2008) and overweight people (Brochu, Gawronski, & Esses, 2011). In addition, we found corresponding patterns for four different measures of implicit prejudice: the standard version of the Implicit Association Test (Greenwald et al., 1998), the personalized Implicit Association Test (Olson & Fazio, 2004), the affect misattribution procedure (Payne, Cheng, Govorun, & Stewart, 2005), and a sequential priming paradigm using a lexical decision task (Wittenbrink, Judd, & Park, 1997).

Figure 18.2 depicts the results of a meta-analysis including the data from all four studies. As predicted by the integrative prejudice framework, implicit and explicit prejudice were positively related when perceptions of discrimination were high and egalitarianism-related, nonprejudicial goals were weak. In addition, implicit and explicit prejudice were positively related when perceptions of discrimination were low and egalitarianism-related, nonprejudicial goals were strong. Interestingly, participants with high levels of perceived discrimination and strong egalitarianism-related, nonprejudicial goals showed a significant negative relation between implicit and explicit prejudice. Drawing on considerations by Dunton and Fazio (1997), this finding may reflect an overcompensation effect, such that more extreme affective responses led these participants to try harder to overcome their negative affective responses by reporting more favorable evaluations on the self-report measure. Consistent with this assumption, negative correlations between direct and indirect measures are in fact rather common whenever cognitive inconsistency leads to a rejection of immediate affective reactions as a basis for an evaluative judgment (e.g., Gawronski & Strack, 2004).

### Discovery of a Previously Undetected Type of Prejudice

Even though the main predictions of our integrative prejudice framework have been generally confirmed, one particular finding was not anticipated on the basis of our framework. Counter to our expectation that participants with low levels of perceived discrimination and weak egalitarianism-related, nonprejudicial goals would also base their evaluative judgments on their immediate affective reactions, these participants showed relatively high levels of explicit prejudice regardless of their levels of implicit prejudice (see Figure 18.2). What seems remarkable about this pattern is that these par-



**FIGURE 18.2.** Meta-analytic regression results ( $N=392$ ) reflecting the relation between negative affective reactions toward stigmatized groups (implicit prejudice) and direct evaluative judgments about these groups (explicit prejudice) as a function of perceived discrimination and egalitarianism-related, nonprejudicial goals. The meta-analytic regression includes data from three studies on prejudice against Black people (Gawronski, Peters, Brochu, & Strack, 2008) and one study on prejudice against overweight people (Brochu, Gawronski, & Esses, 2011) using four different measures of implicit prejudice.

Participants endorsed rather negative evaluations even when they showed low levels of affective negativity. In other words, it seems that these participants exaggerated their experienced negativity to the respective target groups. To our knowledge, this type of prejudice has not yet been identified in the social psychological literature. An important task for future research is to investigate why low levels of perceived discrimination and egalitarianism-related, nonprejudicial goals are associated with an exaggeration of experienced negativity in self-reports.

In evaluating the relation between this unexpected finding and our theoretical framework, a couple of issues deserve further discussion. First, when we developed the integrative prejudice framework, we focused primarily on the three possible cases that involve a rejection of one of three relevant propositions (see Figure 18.1). This focus was guided by the intended integration of old-fashioned, modern, aversive, and implicit prejudice. For these three cases, our predictions have been generally confirmed (see Figure 18.2).

However, these three cases differ from the one producing the unanticipated outcome, in that this case involves a rejection of two rather than one of the involved propositions.

Second, even though we did not anticipate the pattern obtained for participants with low levels of perceived discrimination and weak egalitarianism-related, nonprejudicial goals, it is important to note that this pattern does not violate the notion of cognitive consistency that serves as the theoretical basis of our model. A central assumption of our model is that people aim to avoid cognitive inconsistency in their prejudice-related belief systems. This assumption would be challenged if some people display a system of beliefs that is inconsistent (e.g., if they endorse all three components depicted in Figure 18.1, Panel A). However, this is not the case for participants with low levels of perceived discrimination and weak egalitarianism-related, nonprejudicial goals. After all, the particular beliefs associated with the two components are consistent with any kind of evaluative judgment, be it positive or negative. Thus, even though the obtained exaggeration of negativity was not expected, it does not contradict our core assumption that prejudice-related belief systems are constrained by the principles of cognitive consistency.

Third, our unanticipated finding suggests that additional belief components should be taken into account to provide a better understanding of the conditions under which people rely on their immediate affective reactions when making evaluative judgments about stigmatized groups. Irrespective of the content of these belief components (e.g., judgments of personal responsibility for being in a disadvantaged position), our integrative prejudice framework suggests that these components should be analyzed from a cognitive consistency perspective. This conclusion implies that research in this area should move beyond investigating zero-order correlations between different belief components (e.g., using structural equation models to describe distal and proximal relations) to studying their interactive relations in comprising an internally consistent system of beliefs.

### An Alternative Way to Maintain Consistency

So far, we primarily have focused on cases that involve the rejection of propositions in prejudice-related belief systems (i.e., subtraction of dissonant cognitions; see Festinger, 1957). However, it is important to note that inconsistency can also be resolved in a different way, namely, by searching for an additional proposition that resolves the inconsistency (i.e., addition of consonant cognitions; see Festinger, 1957). In studies on cognitive dissonance, such additional propositions are usually conceptualized as situational explanations for a particular behavior (e.g., Festinger & Carlsmith, 1959). Applied to our integrative prejudice framework, situational explanations may resolve a potential inconsistency among negative evaluations, perceived discrimination, and egalitarianism-related, nonprejudicial goals (see Figure 18.1, Panel A) when these explanations provide a justification for a negative evaluation

despite high levels of perceived discrimination and strong egalitarianism-related, nonprejudicial goals. For instance, people may use their negative affective reaction toward Black people as a basis for a negative judgment when the negative judgment can be legitimized by specific aspects of the situation. Such justification processes play a central role in research on aversive racism (Dovidio & Gaertner, 2004). Confirming a central implication of aversive racism theory, several studies found that White participants' behavior toward Black individuals did not differ from their behavior toward White individuals under control conditions. However, when negative behavior could be attributed to factors other than racial prejudice, participants' behavior toward Black individuals tended to be more negative than their behavior toward White people (e.g., Gaertner, 1973). Our integrative prejudice framework implies the novel, yet untested, prediction that such negative behavior under conditions of attributional ambiguity should be driven by immediate affective reactions. Specifically, immediate affective reactions should influence judgments and behavior when a potential inconsistency among negative affective reactions, nonprejudicial goals, and perceived discrimination can be resolved by means of an additional proposition (e.g., situational justification). However, negative affective reactions should leave judgments and behavior unaffected when inconsistency is resolved by a rejection of negative affective reactions. This prediction goes beyond earlier results showing that measures of implicit prejudice predict spontaneous but not deliberate behavior (e.g., Dovidio, Kawakami, & Gaertner, 2002), in that the relation between implicit prejudice and deliberate behavior should be moderated by attributional ambiguity; that is, measures of implicit prejudice should predict deliberate behavior toward members of the relevant target group when attributional ambiguity is high, but not when it is low.

### Attributions of Prejudice

Despite their impact on research and theorizing in social psychology, the concepts of modern, aversive, and implicit prejudice have not been without controversy. For example, research on modern prejudice has been criticized for inferring prejudice from measures that may simply reflect conservative political opinions (Sniderman & Tetlock, 1986). Similarly, research on implicit prejudice has been criticized for inferring prejudice from measures that may reflect cultural stereotypes rather than "personal animus" (Arkes & Tetlock, 2004). Our integrative prejudice framework offers a new perspective on these debates by specifying the relation between several concepts that have been the subject of controversy. Whereas some of the conclusions implied by our model reiterate earlier criticism, others confirm the significance of the proposed concepts for understanding the dynamics of prejudice-related belief systems.

First, our framework agrees with previous criticism that measures of implicit prejudice (tapping the activation of negative associations) or mea-

asures of modern prejudice (tapping perceptions of discrimination) do not reflect the *endorsement* of negative evaluations of a particular target group. By definition, such an endorsement can be assessed only with measures that involve a direct evaluation of these groups, such as feeling thermometer or likability ratings (De Houwer, 2006). Second, our model also agrees with previous criticism that neither the denial of discrimination nor implicit negativity are sufficient for the endorsement of negative evaluations. After all, low levels of perceived discrimination may not necessarily be associated with direct negative judgments when there is no affective negativity at the implicit level. In addition, high levels of implicit prejudice may not be reflected in direct negative judgments when perceived discrimination is high and, at the same time, egalitarianism-related, nonprejudicial goals are strong. Third, our integrative prejudice framework disagrees with previous criticism by arguing that both implicit prejudice and perceptions of discrimination systematically contribute to the endorsement of negative evaluations of social groups. Specifically, our model implies that negative affective reactions (as assessed by measures of implicit prejudice) directly translate into negative judgments when perceived discrimination is low or egalitarianism-related, nonprejudicial goals are weak. However, negative affective reactions should not translate into negative judgments when perceived discrimination is high and, at the same time, egalitarianism-related, nonprejudicial goals are strong. As outlined earlier, these predictions have been empirically confirmed in several studies using different target groups and multiple measures of implicit prejudice (see Figure 18.2). Thus, theoretical controversies about the nature of different forms of prejudice may be resolved by analyzing the specific relations between the proposed concepts from a cognitive consistency perspective.

### IMPLICATIONS FOR PREJUDICE REDUCTION

In addition to specifying the relation between different forms of prejudice, our framework also has important implications for prejudice reduction (for a more elaborate discussion, see Sritharan & Gawronski, 2010). The most significant implication is that targeting single components of prejudice-related belief systems may often be ineffective in producing the desired changes in overt behavior. For instance, simply increasing the awareness of ongoing social discrimination may be unsuccessful in reducing prejudiced responses when egalitarianism-related, nonprejudicial goals are weak (see Figure 18.1, Panel B). Similarly, successful attempts to enhance egalitarianism-related, nonprejudicial goals may leave prejudiced responses unaffected when the relevant group is not considered to be a target of social discrimination (see Figure 18.1, Panel C).<sup>3</sup> From the perspective of our framework, attempts to reduce prejudice will be most effective if they are simultaneously directed at all of the relevant components. Because associative and propositional

processes often respond differently to the same manipulation (Gawronski & Bodenhausen, 2006), effective prejudice reduction ultimately requires a well-informed combination of different strategies.

### Changing Propositional Beliefs

In technical terms, changes in propositional beliefs can be described as a change in the subjective truth or falsity of a given proposition (Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). Such changes in the ascription of truth values are often due to cognitive inconsistency resulting from a change in the set of considered propositions (e.g., as a result of acquiring new information). One important source of such changes is exposure to persuasive messages. Encoding persuasive messages with counterattitudinal content adds inconsistent components to the current set of beliefs, which in turn may produce changes in propositional beliefs as result of the desire to maintain cognitive consistency (for a more detailed discussion, see Festinger, 1957, Chapter 6). In line with these considerations, prejudice has often been targeted with persuasive campaigns that appeal to people's values and beliefs, with the goal of questioning the morality or legality of prejudice (for a review, see Amodio & Devine, 2005). The general expectation is that by persuading prejudiced individuals to adopt prejudice-inconsistent values or beliefs, these individuals may change their propositional beliefs about stigmatized groups. Applied to our integrative prejudice framework, persuasive messages may be used to increase egalitarianism-related, nonprejudicial goals (e.g., by appealing to personal values) and perceptions of discrimination (e.g., by providing factual information), which represent two of the central components of prejudice-related belief systems (see Figure 18.1).

Another potential source of cognitive inconsistency is counterattitudinal behavior. The most prominent example in this regard is Festinger and Carlsmith's (1959) induced compliance study, showing changes in people's self-reported evaluations as a result of counterattitudinal behavior when situational incentives to engage in that behavior were low, but not when incentives were high. Applying these findings to the domain of prejudice, Leippe and Eisenstadt (1994) found that White participants who were asked to write an essay supporting an increase in scholarships for Black students reported less negative evaluations of Black people when the situational pressure to write the essay was low than when the situational pressure was high. These results suggest that engaging prejudiced individuals in positive behaviors toward stigmatized groups can be helpful in reducing negative judgments about these groups, at least when there is no salient situational explanation for these behaviors. However, from the perspective of our model, it is important to note that this strategy may be ineffective in reducing negative affective reactions. After all, counterattitudinal behavior may simply enforce a change in the subjective truth or falsity of a particular evaluation, without reducing its underlying affective reaction. In line with this contention, Gawronski

and Strack (2004) found that even though Leippe and Eisenstadt's (1994) manipulation effectively reduced people's self-reported negative judgments, implicit prejudice remained unaffected. Moreover, explicit and implicit prejudice showed a positive correlation when situational pressure was high, but a tendency toward a negative correlation when situational pressure was low. These results suggest that standard dissonance manipulations could possibly be used to change the subjective validity of propositional components of prejudice-related belief systems. However, they may be less effective in reducing negative affective reactions.

### Changing Affective Reactions

Drawing on generalized dual-process models (Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004), we argue that people experience positive or negative affective reactions to a given object depending on the evaluative quality of the associations activated in response to that object. From this perspective, changes in affective reactions can be effected by changing the evaluative associations that are responsible for these reactions. Such changes may be further differentiated into (1) momentary changes in the activated pattern of associations or (2) changes in the underlying structure of chronic associations (Gawronski & Bodenhausen, 2006).

Changes caused by variations in pattern activation can occur when contextual cues activate different subsets of the associative representation of an object. For instance, responses to the same individual (e.g., Michael Jordan) may differ as a function of whether this person is categorized in terms of race or an alternative category, such as occupation (Mitchell, Nosek, & Banaji, 2003). Hence, evaluative responses are often context-sensitive, in that contextual cues may influence how a particular individual is categorized (Fazio, 2007). Conversely, there is evidence that cognitively accessible exemplars influence the momentary construal of a social group, which in turn influences immediate affective reactions to that group (e.g., Dasgupta & Greenwald, 2001). However, in evaluating the effectiveness of these mechanisms in reducing prejudice, it is important to note that the resulting changes reflect only momentary variations in activated associations, which can differ as a function of context and time (e.g., Joy-Gaba & Nosek, 2010). As such, they are less suited to bring about long-lasting and context-independent changes in overt behavior.

A strategy that seems more effective in producing stable outcomes is to change the chronic associations that are responsible for enduring affective responses. One manipulation that may produce such changes is the repeated pairing of a social group with positive stimuli. Research on evaluative conditioning (EC) has shown that repeated pairings of a conditioned stimulus (CS) with a positive or negative unconditioned stimulus (US) tend to produce evaluations of the CS that are in line with the valence of the US with which it has been paired (for a review, see De Houwer, Thomas, & Baeyens, 2001).

Attesting to the usefulness of EC procedures in reducing prejudice, Olson and Fazio (2006) found reduced levels of implicit racial prejudice when Black faces were repeatedly paired with positive stimuli, and these effects remained stable when measured 2 days later. Note, however, that despite significant reductions in implicit prejudice, there were no EC-related changes in self-reported evaluative judgments. This dissociation corroborates our concern that targeting single components in prejudice-related belief systems—in this case, evaluative associations—may be less effective in producing generalized changes in overt behavior. Instead, reducing prejudice requires targeting all relevant components of prejudice-related belief systems, not just evaluative associations. After all, the affective reactions resulting from newly created associations may be rejected as a basis for an evaluative judgment, which in turn may produce changes in implicit but not explicit prejudice (see Gawronski & LeBel, 2008).

Another important line of research in this context concerns the differential effectiveness of affirmation versus negation foci in changing evaluative associations. A central assumption of our model is that merely negating the validity of one's affective response to a given object does not necessarily deactivate the associations responsible for the affective response (Deutsch et al., 2006). To the contrary, repeatedly negating an evaluative proposition (e.g., "It is not true that old people are bad drivers") may strengthen the association that underlies that proposition (i.e., the association between *old people* and *bad drivers*), thereby leading to ironic effects on immediate affective responses (Wegner, 1994). From the perspective of our model, the more effective strategy is to affirm the opposite proposition (e.g., "Old people are good drivers"), which strengthens the respective alternative association (i.e., the association between *old people* and *good drivers*). In line with these assumptions, Gawronski, Deutsch, Mbirkou, Seibt, and Strack (2008) found that repeated affirmation of positive associations reduced implicit prejudice, while repeated negation of negative associations led to an increase in implicit prejudice.

### Simultaneously Targeting Propositional Beliefs and Affective Reactions

Our integrative prejudice framework suggests that strategies to reduce prejudice are most effective if they simultaneously target all components of prejudice-related belief systems. Such strategies impose more constraints on a person's belief-system, thereby reducing the number of alternative means to achieve consistency (e.g., compensating the effect of increased egalitarianism-related, nonprejudicial goals by denying continued discrimination). Inconsistency can be resolved in multiple ways (Festinger, 1957) and imposing multiple constraints leaves less room for influences of motivated reasoning (Kunda, 1990). A useful example in this regard is a diversity seminar on

prejudice and intergroup conflict designed and evaluated by Rudman, Ashmore, and Gary (2001). These researchers found that participation in their seminar reduced both explicit and implicit prejudice. In line with the implications of our integrative prejudice framework, changes in discrimination awareness and egalitarianism-related, nonprejudicial goals were uniquely related to changes in explicit (but not implicit) prejudice, whereas changes in the liking of the course and the Black professor were uniquely related to changes in implicit (but not explicit) prejudice. These results suggest that explicit prejudice reduction requires cognitive change through the stimulation of egalitarianism-related, nonprejudicial goals and by increasing awareness of ongoing discrimination, whereas implicit prejudice reduction may require affective change through a decrease in fear of the target group and positive contact with members of that group (Sritharan & Gawronski, 2010).

## CONCLUSIONS

The main tenet of this chapter is that research on prejudice can benefit from an application of consistency principles in various ways. First, describing prejudice-related belief systems from a cognitive consistency perspective provides a conceptual integration of old-fashioned, modern, aversive, and implicit prejudice in a single theoretical framework (*integration*). Second, this framework implies novel predictions about the relation between implicit and explicit prejudice (*prediction*). Third, empirical tests of these predictions have led to the discovery of a previously undetected form of prejudice (*discovery*). Fourth, the proposed framework can help to increase the effectiveness of prejudice reduction strategies by clarifying the dynamic processes underlying changes in prejudice-related belief systems (*application*). Finally, the principles identified by the proposed framework are applicable to prejudiced beliefs about all kinds of stigmatized groups, thereby meeting the request for general theories that apply to a broad range of phenomena (*generalization*). On the basis of these conclusions, we hope that researchers will appreciate the value of consistency principles in providing a better understanding of prejudice-related belief systems.

## NOTES

1. In line with suggestions by De Houwer (2006), we use the terms *direct* and *indirect* to describe characteristics of measurement procedures, and the terms *explicit* and *implicit* to refer to the constructs assessed by different kinds of measurement procedures.
2. Note that egalitarianism-related, nonprejudicial goals can be conceptualized as a negative evaluation of discriminatory behavior. Thus, perceived discrimination and egalitarianism-related, nonprejudicial goals reflect what we described as non-

evaluative beliefs referring to states of affairs in the world and evaluative beliefs about other attitude objects (Jones & Gerard, 1967).

3. In this context, it is worth noting that the election of Barack Obama as the first African American President of the United States may have had paradoxical effects. Specifically, it seems likely that his election reduced perceptions of racial discrimination, which can legitimize the endorsement of negative evaluations of African Americans (see Figure 18.1, Panel C; for related evidence, see Effron, Cameron, & Monin, 2009; Kaiser, Drury, Spalding, Cheryan, & O'Brien, 2009).

## REFERENCES

- Amodio, D. M., & Devine, P. G. (2005). Changing prejudice: The effects of persuasion on implicit and explicit forms of race bias. In T. C. Brock & M. C. Green (Eds.), *Persuasion: Psychological insights and perspectives* (2nd ed., pp. 249–280). Thousand Oaks, CA: Sage.
- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or “Would Jesse Jackson ‘fail’ the Implicit Association Test? *Psychological Inquiry*, 15, 257–278.
- Brochu, P. M., Gawronski, B., & Esses, V. M. (2008). Cognitive consistency and the relation between implicit and explicit prejudice: Reconceptualizing old-fashioned, modern, and aversive prejudice. In M. A. Morrison & T. G. Morrison (Eds.), *The psychology of modern prejudice* (pp. 27–50). Hauppauge, NY: Nova Science Publishers.
- Brochu, P. M., Gawronski, B., & Esses, V. M. (2011). The integrative prejudice framework and different forms of weight prejudice: An analysis and expansion. *Group Processes and Intergroup Relations*, 14, 429–444.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81, 800–814.
- De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11–28). Thousand Oaks, CA: Sage.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853–869.
- Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, 91, 385–405.
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 1–52). San Diego: Academic Press.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62–68.
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23, 316–326.
- Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of Experimental Social Psychology*, 45, 590–593.
- Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, 25, 603–637.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203–210.
- Franco, F. M., & Maass, A. (1999). Intentional control over prejudice: When the choice of the measure matters. *European Journal of Social Psychology*, 29, 469–477.
- Gaertner, S. L. (1973). Helping behavior and discrimination among liberals and conservatives. *Journal of Personality and Social Psychology*, 25, 335–341.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Orlando, FL: Academic Press.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692–731.
- Gawronski, B., Deutsch, R., Mbirikou, S., Seibt, B., & Strack, F. (2008). When “just say no” is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, 44, 370–377.
- Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, 44, 1355–1361.
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us?: Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181–193.
- Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin*, 34, 648–665.
- Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, 40, 535–542.
- Greeley, A. M., & Sheatsley, P. B. (1971). Attitudes toward racial integration. *Scientific American*, 225, 13–19.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Jones, E. E., & Gerard, H. B. (1967). *Foundations of social psychology*. New York: Wiley.
- Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology*, 41, 137–146.
- Kaiser, C. R., Drury, B. J., Spalding, K. E., Cheryan, S., & O'Brien, L. T. (2009). The ironic consequences of Obama's election: Decreased support for social justice. *Journal of Experimental Social Psychology*, 45, 556–559.
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Cor-

- relational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology*, 55, 893–905.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498.
- Leippe, M. R., & Eisenstadt, D. (1994). Generalization of dissonance reduction: Decreasing prejudice through induced compliance. *Journal of Personality and Social Psychology*, 67, 395–413.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. D. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic Press.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132, 455–469.
- Nail, P. R., Harton, H. C., & Decker, B. P. (2003). Political orientation and modern versus aversive racism: Tests of Dovidio and Gaertner's (1998) integrated model. *Journal of Personality and Social Psychology*, 84, 754–770.
- Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, 86, 653–667.
- Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, 32, 421–433.
- Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277–293.
- Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, 81, 856–868.
- Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, 17, 437–465.
- Schuhman, H., Steeh, C., Bobo, L., & Kyrsan, L. (1997). *Racial attitudes in America: Trends and interpretations*. Cambridge, MA: Harvard University Press.
- Sears, D. O. (1988). Symbolic racism. In P. A. Katz & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 53–84). New York: Plenum.
- Sniderman, P. M., & Carmines, E. G. (1997). Reaching beyond race. *Political Science and Politics*, 30, 466–471.
- Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues*, 42, 129–150.
- Son Hing, L. S., Chung-Yan, G., Grunfeld, R., Robichaud, L., & Zanna, M. P. (2005). Exploring the discrepancy between implicit and explicit prejudice: A test of aversive racism theory. In J. P. Forgas, K. Williams, & S. Latham (Eds.), *Social motivation: Conscious and unconscious processes* (pp. 274–293). New York: Psychology Press.
- Son Hing, L. S., Chung-Yan, G. A., Hamilton, L. K., & Zanna, M. P. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology*, 94, 971–987.
- Sritharan, R., & Gawronski, B. (2010). Changing implicit and explicit prejudice: Insights from the associative-propositional evaluation model. *Social Psychology*, 41, 113–123.

- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247.
- Taylor, D. G., Sheatsley, P. B., & Greeley, A. M. (1978). Attitudes toward racial integration. *Scientific American*, 238, 42–49.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, 101, 34–52.
- Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationships with questionnaire measures. *Journal of Personality and Social Psychology*, 72, 262–274.
- Zarate, M. A. (2009). Racism in the 21st century. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 387–406). New York: Psychology Press.