being endorsed by a number of legal writers. Despite these writers and the *Jackson* decision, however, attorneys frequently challenge fact finders explicitly to resist succumbing to bias in making judgments about members of stereotyped groups. For example, in the recent World Trade Center bombing case, defense attorney Austin Campriello asked the jury to avoid associating stereotypes of Arab and Muslim violence and terrorism with his client. And in a recent capital murder trial, defense attorney Paul Nugent urged the jury not to allow homophobia to distort their deliberations about his client’s guilt or innocence.

Thus we have two radically conflicting strategies for helping decision makers achieve greater objectivity: color-blindness versus color-consciousness. Do arguments based on race, sexual preference, or any other characteristic widely used to stereotype individuals necessarily “appeal to prejudice,” or instead can some such arguments actually promote the rationality and fairness of the fact-finding process? Does race-consciousness help decision makers avoid imposing the Black tax on African Americans, or does it hike up the tax? Are proactive measures for reducing private discrimination feasible, or must we resign ourselves to the inevitability of biased social judgments? We shall begin our discussion with this last question, for a number of thoughtful and progressive thinkers and activists maintain that prejudice is inevitable. If this rather fatalistic view is correct, we truly shall remain lost in a wilderness of racial injustice, and the dream of Martin Luther King, Jr., shall remain forever deferred.

**Hypocritical Racists and Aversive Racists**

Americans share a cultural belief system saturated with derogatory stereotypes about Blacks; thus we have all been influenced
by these stereotypes. Does this mean that we are all racists? Leading thinkers in law, psychology, and social science answer this question with an emphatic “yes.” They view prejudice as an inevitable outgrowth of our stereotype-ridden cultural belief system.7

This view of prejudice as inevitable and ubiquitous, however, does not square with polls and studies indicating that prejudice has been declining steadily over the past forty years. Between 1956 and 1978, reports on attitudes of White Americans toward Black Americans show a steady increase in the percentage of Whites who favor equality for Blacks in all areas of American society.8 The most common objection to these studies is that they do not capture the resurgence of racism in the late 1980s. A review of studies and surveys conducted between 1984 and 1990 on young White adults, however, showed that there was no significant decline in liberal racial attitudes among men and women who became adults between 1960 and 1990.9

Progressives who maintain that we are all racists in spite of these studies attack the validity of the studies. Because the studies rely on the self-evaluations of the people being polled, critics in effect characterize persons who report nonprejudiced personal beliefs as either hypocritical, sub rosa racists or unconscious, aversive racists.

HYPOCRITICAL RACISTS

In Black Innocence and the White Jury, Sheri Lynn Johnson, a noted expert on racism in criminal justice, sees hypocrisy playing a big role in the self-evaluations that seem to indicate growing racial tolerance.10 According to Johnson, “[A]ny encouragement that might be drawn from the initial decrease in extreme negative stereotypes must be qualified by the likelihood that
newer data reflect some fading of stereotypes—but also some faking.” From this viewpoint, prejudice has not decreased nearly as much as it seems; it has just become less socially acceptable. Thus, merely to appear socially desirable, many survey respondents profess racial liberalism. Although Johnson does not give a concrete estimate of how much “faking” the newer data reflect, she does suggest that “it now may be quite common to underreport prejudiced attitudes” by faking racial tolerance.

To support the hypocrisy interpretation, Johnson points to the findings of an experiment in which White subjects were asked to report their responses to Blacks under a normal (control) condition and under a “bogus pipeline” condition. In the pipeline condition, a researcher wires his subjects to a machine that the subjects believe will give him an accurate physiological measure of (i.e., a pipeline to) their automatic or “covert” reactions. The researchers then asked these subjects to estimate what the machine was telling the experimenter about their uncontrolled responses to Blacks, as the experimenter asked them to rate Blacks on various personality traits, such as ignorance, stupidity, honesty, and sensitivity. Researchers assumed that these estimations would correspond to the subjects’ “honest” beliefs about Blacks. Subjects’ estimates of their uncontrolled responses to Blacks in the pipeline condition were significantly more negative than the responses to Blacks reported by subjects who did not believe that their uncontrolled responses were being monitored.

Johnson characterizes these automatic, uncontrolled physiological responses in this experiment as the subjects’ “true feelings” and “pure” attitudes. She interprets these findings as proof of prejudice’s persistence notwithstanding survey data to the contrary. This interpretation, however, rests on a failure to
distinguish between two distinct sources of negative responses to Blacks (and other marginalized social groups)—namely, stereotypes and prejudice. Once this critical distinction is understood, it becomes evident that the bogus pipeline results prove only the persistence of stereotypes, not prejudice, and therefore are perfectly consistent with the proposition that prejudice has decreased significantly over the last forty years.

Stereotypes consist of well-learned sets of associations among groups and traits established in children’s memories at an early age, before they have the cognitive skills to decide rationally upon the personal acceptability of the stereotypes. For example, Dr. Phyllis Katz reports a chilling case of a three-year-old child who, upon seeing a Black infant, said to her mother, “Look, Mom, a baby maid.” By the time the child turned three, before she had developed the cognitive ability to judge the appropriateness of the stereotypic ascription, the associational link between Black women and certain social roles was already forged in her memory.

In contrast, prejudice consists of derogatory personal beliefs. “Beliefs” are propositions that people endorse and accept as being true. Thus, prejudiced personal beliefs are the endorsement or acceptance of a negative cultural stereotype. That a person has a negative stereotype established in her memory does not necessarily mean that she endorses that stereotype. As Patricia Devine, a leading psychologist who has won prestigious awards for her trailblazing research in this area, points out, “Although one may have knowledge of a stereotype, his or her personal beliefs may or may not be congruent with the stereotype.” For example, if the three-year-old child described above grows up and decides that the stereotype of a maid is an inappropriate basis for responding to Black women, she may
experience a fundamental conflict between the previously established stereotype and the more recently established non-prejudiced personal belief. In such a case, her responses to Black women and to Blacks generally will turn on whether those responses are based on the well-established stereotype or her more recently adopted nonprejudiced beliefs.

Of course, some people’s stereotypes and personal beliefs overlap; that is, some people not only have knowledge of the cultural stereotypes from years of socialization, but they endorse and accept them as well. We shall refer to these individuals as high prejudiced. However, many people have thought about the cultural stereotypes, recognized them as inappropriate bases for responding to others, and deliberately rejected them. We shall refer to these individuals as low prejudiced. Although high- and low-prejudiced persons differ in their personal beliefs about Blacks, common socialization experiences have firmly entrenched the cultural stereotype of Blacks in the memories of both.

The failure to distinguish between stereotypes and prejudiced personal beliefs leads Johnson and other commentators to take an all-or-none approach to prejudice: if a person experiences any stereotype-congruent responses in any situation, she is prejudiced. This view fails to recognize that a change in a person’s beliefs does not instantly extinguish habitual responses derived from well-learned stereotypes. Because stereotypes are established in children’s memories at an early age and constantly reinforced through the mass media and other socializing agents, stereotype-congruent responses may persist long after a person has sincerely renounced prejudice. Nonprejudiced beliefs and stereotype-congruent thoughts and feelings may coexist within the same individual. Dr. Thomas Pettigrew, a leading
authority on stereotypes and prejudice, has described one example of this conflict: “Many Southerners have confessed to me . . . that even though in their minds they no longer feel prejudice against Blacks, they still feel squeamish when they shake hands with a Black. The feelings are left over from what they learned in their families as children.”

There is strong empirical evidence that the vast majority of low-prejudiced people realize that they are prone to stereotype-congruent responses, that is, that their *actual* reactions to Blacks and other socially marginalized groups sometimes conflict with their personal standards for how they *should* respond. In one recent study, researchers gave a sample of several hundred White subjects (college students, very few of whom were high prejudiced) a questionnaire, the first section of which asked them to report their personal standards for how they *should* respond in five different situations involving Black people. For example, one situation read as follows: “Imagine that a Black person boarded a bus and sat next to you. You *should* feel uncomfortable that a Black is sitting next to you.”

The subjects were asked to circle the number between 1 (strongly disagree) and 7 (strongly agree) that best reflected their personal standard for how they *should* respond in each situation. The second section of the questionnaire asked the subjects to report on the 1-to-7 scale how they believed they actually *would* respond in the same five situations. Out of the 101 cases, 71 percent of the subjects reported actual *would* responses that were more negative than their *should* responses, which reflected their personal standards for how they should respond. Separate studies found similar should-would discrepancies in responses to homosexual men. These studies also investigated whether the subjects’ personal standards (shoulds)
were well internalized (i.e., viewed by the subjects as highly important and as central to their personal identity or merely derived from society’s standards). Researchers found that low-prejudiced subjects strongly internalized their personal standards, and that these subjects felt compunction (guilt and self-criticism) when they transgressed the standards.31

These findings, which several later studies have replicated, suggest a less-pessimistic interpretation of the bogus pipeline results than Johnson adopts. That a subject reports more negative responses about Blacks when he believes an experimenter can monitor his autonomic nervous system (or what Johnson refers to as his “true attitudes”) does not prove that he is truly prejudiced or that he is faking his more positive responses on questionnaires. It may show only that he realizes, as most low-prejudiced people do, that he is prone to stereotype-congruent responses. Although he may not endorse these responses and may feel compunction about experiencing them, he may believe that the pipeline will detect their presence. In other situations, however, such as responding to a questionnaire, the low-prejudiced person may inhibit his stereotype-congruent responses and replace them with responses based on his nonprejudiced personal standards. A model of prejudice that recognizes the distinction between stereotypes and prejudiced personal beliefs—a model we shall call the “dissociation model”—points to the possibility of inhibiting and replacing stereotype-congruent responses with nonprejudiced responses derived from nonprejudiced personal beliefs. If nonprejudiced personal beliefs can counteract stereotypes in this way, perhaps there is hope for combating the influence of ubiquitous derogatory stereotypes.

Responses derived from nonprejudiced personal beliefs can inhibit and replace responses derived from stereotypes. Low-
and high-prejudiced people, as discussed below, are equally prone to stereotype-congruent responses when they cannot consciously monitor their responses to questions. However, low- and high-prejudiced people have given very different responses when they have had to think consciously about what their responses imply about their self-image. For example, one study asked subjects to list all of their own thoughts (e.g., beliefs, feelings, expectancies) about Blacks under strictly anonymous conditions, thus eliminating any reason to manufacture “correct” responses. The high-prejudiced subjects listed primarily negative stereotypical thoughts about Blacks and were inclined to stereotype. In contrast, the low-prejudiced subjects wrote few pejorative thoughts; they reported beliefs that contradicted the stereotype and emphasized the importance of racial equality. These results make intuitive sense. For low-prejudiced people, writing stereotype-congruent thoughts would contradict their personal beliefs and threaten their non-prejudiced identity. But because beliefs of high-prejudiced people overlap with stereotypes, conscious reflection should not inhibit their stereotype-congruent responses. Thus, if personal beliefs really matter, if they can counteract the stereotype-congruent responses to which research shows high- and low-prejudiced people are equally prone, then the thoughts that low-prejudiced subjects anonymously list about Blacks should be very different from the thoughts anonymously listed by high-prejudiced subjects.

The findings of this thought-listing study, which are strongly confirmed and extended by other research discussed below, reveal that much more than semantics is at stake in the distinction between stereotypes and prejudice. For inasmuch as negative stereotypes and personal beliefs diverge, as they do in low-
prejudiced people, they imply different responses to stereotyped groups. This insight enables us to investigate the interplay between the two conceptually distinct sets of responses, and to develop strategies for activating the responses based on non-prejudiced personal beliefs and inhibiting the stereotype-congruent responses. Before elaborating a framework for working out the full implications of the interplay between stereotypes and non-prejudiced personal beliefs, we must consider the other major attack on the validity and efficacy of non-prejudiced personal beliefs: aversive racism.

**AVERSIVE RACISTS**

A dominant model of prejudice is the theory of aversive racism. Whereas the hypocritical racist model posits that people who express non-prejudiced personal beliefs are manipulating their self-presentation to appear more socially desirable, the aversive racist model holds that ostensibly non-prejudiced people are not so much deceiving others as fooling themselves. The two models are not mutually exclusive, but complementary. Progressive critics freely switch from one to the other in dismissing the validity of people’s racially liberal self-descriptions and non-prejudiced personal beliefs.

The theory of aversive racism begins with the proposition that most Americans are highly committed to egalitarian values. Therefore, they desire to maintain an egalitarian, non-prejudiced self-image. This desire causes them to express non-prejudiced personal beliefs. Such professed non-prejudiced beliefs are not to be confused with genuine—that is, well-internalized—non-prejudiced beliefs, for deep down “aversive racist[s] believe[] in White superiority” and “do not want to associate with Blacks.” Desperately clinging to their egalitarian, non-
prejudiced values and self-image, aversive racists repress their negative feelings and beliefs about Blacks. These repressed anti-Black beliefs have been called “hidden prejudice.” Because aversive racists do not recognize their anti-Black attitudes, the prospects for prejudice reduction are particularly dim. Here the pessimism of the aversive racism model asserts itself. Writing from this perspective, one progressive writer observes, “It is difficult to change an attitude that is unacknowledged. Thus, ‘like a virus that mutates into new forms, old-fashioned prejudice seems to have evolved into a new type that is, at least temporarily, resistant to traditional . . . remedies.’”

Proof of aversive racism, some believe, lies in the discrepancy between responses to Blacks that are consciously monitored and those that are not consciously monitored. Whenever aversive racists consciously monitor their responses to Blacks, they do not discriminate against them since discrimination would undermine their egalitarian self-images. For example, verbal responses to questionnaires designed to measure racial prejudice can be monitored consciously by the respondents and therefore cannot identify aversive racists. More generally, if the situation clearly calls for a nonprejudiced response or if a nonracial justification or rationalization for engaging in a prejudiced response cannot be generated, the response will be positive because it cannot escape being consciously monitored.

In contrast, when the situation is normatively ambiguous, or when a nonrace-related justification is handy, the covert anti-Black attitudes and beliefs of aversive racists find expression in racial discrimination. For example, White research subjects led to believe that a person was in distress helped Black victims as often as White victims when there was no ostensible justification for a failure to help. However, if the subjects knew that someone was
available to help, they “helped Black victims much less frequently than they helped White victims (38% vs. 75%).”42 According to proponents of the aversive racism model, the availability of other potential rescuers provided subjects with a convenient nonracial excuse for not helping the Black victims. This interpretation of the helping behavior study carries very discouraging implications for racially fair dispute resolution. For in courtrooms, for example, finding a nonracial reason to discriminate against a Black litigant is especially easy to do—one simply gives more weight to the evidence favoring the opposing litigant.

The aversive racism model, however, is empirically and conceptually incomplete. One empirical problem with the model stems from its assumption that aversive racists—who, according to commentators, now include most Americans—are not aware of their conflicting reactions to Blacks; their anti-Black thoughts and feelings are supposedly excluded from consciousness. If this assumption were accurate, most survey respondents would not report discrepancies between their standards for how they should respond to Blacks and how they actually would respond, since they are unaware. Yet the vast majority of subjects in several studies recognized and acknowledged that they sometimes experience such discrepancies. Thus, although the aversive racism framework may describe some White Americans, it almost certainly does not account for most.

Another empirical problem with the aversive racism theory concerns the Freudian theory of unconscious motivation to which it is often wedded.43 Psychoanalytic theory presents real difficulties for empirical verification. A model, such as aversive racism, whose theoretical underpinnings are not empirically demonstrable demands an intellectual leap of faith that many may be unwilling to make.
The conceptual problem with the aversive racism model grows out of its tendency to conflate stereotype and prejudice. One implication of aversive racism theory—an implication explicitly embraced by some of its supporters—is that because we have all been influenced by derogatory information about Blacks, we are all racists. This conclusion fails to fully consider that people do not always endorse the knowledge structures that socialization has established in their memories. For example, although socializing forces undoubtedly have entrenched the cultural stereotype of women in the memory of feminists as well as every other American, feminists could be called “sexists” only in a Pickwickian sense. One reason it seems so anomalous to apply the value-laden term “sexist” to feminists is because feminists have both renounced the cultural stereotype about women and developed egalitarian personal beliefs about women. Thus, feminists have two distinct and conflicting cognitive structures concerning women: the cultural stereotype and their egalitarian personal beliefs. Similarly, low-prejudiced people have two conflicting cognitive structures concerning Blacks: the Black cultural stereotype and their nonprejudiced personal beliefs. Calling feminists “sexists” and low-prejudiced persons “racists” identifies them more with the well-learned cultural stereotype than with their personal beliefs, and implies that the stereotype is somehow the more compelling of the two knowledge structures.

What we need is a new, empirically demonstrated, conceptually coherent theory of prejudice. The theory must show how unconscious discrimination drives the social judgments of all Americans. Having diagnosed the ulcer of ubiquitous unconscious bias, it will not do simply to roughly finger it with bitter invective. The theory must also point to proven strategies for