
VIRGINIA LAW REVIEW

VOLUME 95

DECEMBER 2009

NUMBER 8

ARTICLES

MAKING GOOD ON GOOD INTENTIONS: THE CRITICAL ROLE OF MOTIVATION IN REDUCING IMPLICIT WORKPLACE DISCRIMINATION

*Katharine T. Bartlett**

*D*ISCRIMINATION in today's workplace is largely implicit, making it ambiguous and often very difficult to prove. Employment discrimination scholars have proposed reforms of Title VII to make implicit discrimination easier to establish in court and to expand the kinds of situations to which liability attaches. The reform proposals reflect a broad consensus that strong legal norms are crucial to addressing the problem. Yet it is mistaken to assume that strengthening plaintiffs' hands in implicit discrimination cases will necessarily achieve the long-term goal of reducing its occurrence. This Article brings together several strands of social science research showing that (1) implicit bias is not only invisible and largely unintended, but not readily reachable through legal coercion; (2) people whose motivation to act in nondiscriminatory ways is based on an

* A. Kenneth Pye Professor of Law, Duke University School of Law. I am grateful for the communities of scholars at New York University School of Law and Columbia Law School where I did most of the research for this Article on my 2007–2008 sabbatical visits. I also appreciate the opportunities to present this work in its early stages at workshops at Columbia Law School, Duke Law School, and the University of Kentucky School of Law, and at the Graylyn Lecture at Wake University School of Law when the work was near completion. Special thanks for the helpful comments I received along the way from Stuart Benjamin, James Boyle, Erwin Chemerinsky, Justin Driver, Elizabeth Emens, Catherine Fisk, Suzanne Goldberg, Mitu Gulati, Lewis Kornhauser, Kimberly Krawiec, Martha Minow, Jed Purdy, Elizabeth Scott, Chris Schroeder, Neil Siegel, Susan Sturm, and Neil Vidmar, and for research assistance from James McKell and Jennifer McGinnis.

internal commitment to nondiscriminatory norms—or “good intentions”—are less likely to engage in stereotyping of others than people who feel pressured by the law; (3) people internalize nondiscrimination values best when they feel a sense of autonomy, competence, and relatedness; (4) the conditions that support these characteristics in the workplace include strong, unambiguous norms, trust, teamwork, leadership, positive example, and opportunities to grow and advance; and (5) excessive legal control and pressure undermine people’s sense of autonomy, competence, and relatedness and thus their commitment to nondiscrimination norms.

When legal pressure becomes overkill is not a matter of exact science, and is complicated by differences among people and workplace cultures. Still, before fashioning further legal tools that assume that more coercion is the answer to implicit discrimination, this Article suggests that more attention be given to the negative impact of such tools and to alternative measures that may better motivate people’s adherence to nondiscrimination norms.

INTRODUCTION.....	1895
I. THE PSYCHOLOGICAL DIMENSIONS OF RACE AND GENDER DISCRIMINATION.....	1904
A. <i>Stereotyping as Social Cognition</i>	1908
B. <i>Stereotyping as Motivated Prejudice</i>	1911
II. TITLE VII AND IMPLICIT BIAS	1920
A. <i>Title VII Requires Causation, Not Deliberate Intent</i>	1920
B. <i>Proposed Reforms</i>	1926
III. PEOPLE’S MOTIVATIONS TO COMPLY WITH NONDISCRIMINATION NORMS.....	1930
A. <i>Distinguishing External and Internal Motivation</i>	1931
B. <i>The Role of Law in Building, or Dissipating, Internal Motivation to Reduce Implicit Bias</i>	1934
1. <i>The Positive Role of Law on Norm Internalization</i>	1934
2. <i>The Negative Role of Law on Norm Internalization</i>	1936
IV. STRATEGIES TO REDUCE IMPLICIT BIAS.....	1941
A. <i>Effort</i>	1942
B. <i>Exposure and Contact</i>	1947
1. <i>Counterstereotypes</i>	1947

2009]	<i>Making Good on Good Intentions</i>	1895
	2. <i>Minimal Contact and Peer Influence</i>	1950
	3. <i>Working Relationships and the “Contact Hypothesis”</i>	1953
V.	THE IMPLICATIONS OF MOTIVATION ANALYSIS FOR LEGAL AND STRUCTURAL REFORM	1956
	A. <i>Legal Reform Proposals</i>	1956
	B. <i>Structural Reform Proposals</i>	1960
	CONCLUSION.....	1971

INTRODUCTION

A developing body of social psychology research suggests that race and gender bias in this society¹ is invisible, deep, and pervasive,² and sometimes leads to discriminatory

¹This Article addresses discrimination based primarily on race and gender. Examination of the differences between race discrimination and gender discrimination are beyond the scope of this Article, although the data suggest that differences do exist. See, e.g., Alexander M. Czopp & Margo J. Monteith, *Confronting Prejudice (Literally): Reactions to Confrontations of Racial and Gender Bias*, 29 *Personality & Soc. Psychol. Bull.* 532, 541–42 (2003) (finding that people feel more guilty and upset about offending blacks than women). This Article draws upon some social psychology research relating to stereotyping and bias based on age, sexual orientation, and disability, although there are differences here as well that this Article does not examine.

²For a review of the research on implicit bias, see Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 *Cal. L. Rev.* 945 (2006). Much of the research for such conclusions is drawn from the Implicit Association Test (IAT), which is a computer-based test that measures time-response differentials to positive and negative associations related to race, sex, disability, and a number of other stereotype-ridden characteristics. For a description of other tests developed to measure implicit bias, which include not only time-response measure but also word-fragment completion drills, priming tasks, sentence-completion tasks, and cardiovascular measures, see Russell H. Fazio & Michael A. Olson, *Implicit Measures in Social Cognition Research: Their Meaning and Use*, 54 *Ann. Rev. Psychol.* 297, 299–300 (2003).

On the basis of the IAT, Jerry Kang and Mahzarin Banaji make the “conservative estimate” that “seventy-five percent of Whites (and fifty percent of Blacks) show anti-Black bias, and seventy-five percent of men and women” associate female with family more easily than they do with career. Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of “Affirmative Action,”* 94 *Cal. L. Rev.* 1063, 1072 (2006); see also Brian A. Nosek et al., *Harvesting Implicit Group Attitudes and Beliefs From a Demonstration Web Site*, 6 *Group Dynamics: Theory, Res., & Prac.* 101, 112 (2002) (finding that IAT research indicates that all social groups hold implicit biases, regardless of age, gender, race, and political views).

behavior.³ While some scholars criticize the methodology and conclusions of the research,⁴ there can be little doubt that people do act in unconsciously discriminatory ways and that this discrimination is very difficult to regulate. For example, people often respond to members of other groups with lack of eye contact and warmth,

³ See Anthony M. Greenwald & Mahzarin Banaji, *Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes*, 102 *Psychol. Rev.* 4, 7 (1995); see also Greenwald & Krieger, *supra* note 2, at 961 (evidence that implicit bias leads to discriminatory behavior is “already substantial”); Anthony M. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 *J. Personality & Soc. Psychol.* 17 (2009) (meta-analysis of IAT research suggests that IAT predicts discriminatory behavior); Jerry Kang, *Trojan Horses of Race*, 118 *Harv. L. Rev.* 1489, 1514 (2005) (“persuasive evidence” that implicit bias predicts disparate behavior).

⁴ For a sampling of the critiques, see Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or “Would Jesse Jackson ‘Fail’ the Implicit Association Test?”*, 15 *Psychol. Inquiry* 257 (2004) (implicit bias research measures cultural stereotypes rather than personal animus, and results can be explained by factors other than prejudice); Hart Blanton & James Jaccard, *Unconscious Racism: A Concept in Pursuit of a Measure*, 34 *Ann. Rev. Soc.* 277 (2008) (evidence suggests that people sometimes lack control over the cause and consequences of their racial biases, but not that they are unconsciously racist); Hart Blanton et al., *Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT*, 94 *J. Applied Psychol.* 567 (2009) (IAT results do not permit predictions of individual-level behaviors, and when various methodological factors are taken into account, are more consistent with pro-black than anti-black bias); Miguel C. Brendl et al., *How Do Indirect Measures of Evaluation Work? Evaluating the Inference of Prejudice in the Implicit Association Test*, 81 *J. Personality & Soc. Psychol.* 760 (2001) (results on IAT test can be explained by factors other than implicit prejudice); Andrew Karpinski & James L. Hilton, *Attitudes and the Implicit Association Test*, 81 *J. Personality & Soc. Psychol.* 774 (2001) (IAT reflects exposure to stereotypes but not endorsement thereof); Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 *Ohio St. L.J.* 1023 (2006) (implicit prejudice research lacks construct and content validity, ignores high error rates and alternative explanations for alleged discriminatory behavior, and draws naïve conclusions about link between laboratory results and the real world).

Philip Tetlock and Gregory Mitchell see the debate between IAT proponents and critics as part of a long-running impasse between a point of view that sees prejudice as ubiquitous and thinks it is the government’s role to stop it (which they call the “statist-intervention” position), and an approach that believes in the power of competition to eliminate irrational biases (which they call the “market-purist” position). Tetlock and Mitchell conclude that both sides should back off of their non-falsifiable propositions and join in “adversarial collaboration” based on shared standards of proof. See Philip E. Tetlock & Gregory Mitchell, *Implicit Bias and Accountability Systems: What Must Organizations Do to Prevent Discrimination?*, 28 *Research in Organizational Behavior* (Barry Straw & Arthur Brief, eds., forthcoming 2009), available at <http://www.law.virginia.edu/pdf/faculty/ImplicitBiasinOrganizationsandAdversarialCollaboration.pdf>.

2009]

Making Good on Good Intentions

1897

tensing of facial muscles, increased blinking, anxious voice tone, embarrassing slips of the tongue, awkward social interactions, and maintenance of physical distance and formality.⁵ These behaviors can lower the performance level of those whose presence has provoked them,⁶ and yet these behaviors are too minor and too common to constitute the type of employment action that could be reasonably banned by nondiscrimination law. Another example is that people form impressions about others based on unconscious stereotypes, which color what information about them is noticed and remembered, and how it is evaluated;⁷ typically, however, at no discernible point in this chain of accumulated impressions is discrimination readily apparent or provable. Further complicating the matter is that people often perceive differently the experiences and events on which discrimination claims are based. They can participate in the same allegedly discriminatory incident, each confident about what they saw, but not agree just what happened.⁸ Given

⁵ See Jennifer Crocker et al., Social Stigma, *in* Handbook of Social Psychology 504, 513 (Daniel T. Gilbert et al. eds., 4th ed. 1998); John F. Dovidio et al., Implicit and Explicit Prejudice and Interracial Interaction, 82 *J. Personality & Soc. Psychol.* 62, 63 (2002); Susan T. Fiske, What We Know About the Problem of the Century: Lessons from Social Science to the Law, and Back, *in* Handbook of Employment Discrimination Research: Rights and Realities 59, 60, 63 (Laura Beth Nielsen & Robert L. Nelson eds., 2005).

⁶ The classic work on this subject is Carl O. Word, The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction, 10 *J. Exper. Soc. Psychol.* 109, 119 (1974) (negative nonverbal behaviors produced poorer performance in job interview setting among both whites and blacks). See also Claude M. Steele & Joshua Aronson, Stereotype Threat and the Intellectual Test Performance of African Americans, 69 *J. Personality & Soc. Psychol.* 797, 805–06 (1995) (subtle cues that remind ethnic minority students of their stigmatized status undermines achievement on academic tests).

⁷ See Linda Hamilton Krieger, The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity, 47 *Stan. L. Rev.* 1161, 1167, 1209, 1213 (1995).

⁸ See generally Samuel L. Gaertner & John F. Dovidio, Understanding and Addressing Contemporary Racism: From Aversive Racism to the Common Ingroup Identity Model, 61 *J. Soc. Issues* 615, 625 (2005) (describing research revealing that whites and blacks perceive of the same encounters in different ways, without being aware of their different perceptions); Russell K. Robinson, Perceptual Segregation, 108 *Colum. L. Rev.* 1093 (2008). In the criminal context, see Dan M. Kahan et al., Whose Eyes Are You Going to Believe?: *Scott v. Harris* and the Perils of Cognitive Illiberalism, 122 *Harv. L. Rev.* 838 (2009) (describing psychological disposition of individuals to observe and resolve disputed facts in a manner supportive of their group identities); Justin D. Levinson, Forgotten Racial Equality: Implicit Bias, Decision-

these kinds of impediments to proving discrimination that is unconscious or implicit, it is small wonder that employment discrimination plaintiffs are far less likely than other civil plaintiffs to win their cases,⁹ and that, faced with the low prospect of success, federal employment discrimination case filings and terminations have dropped almost forty percent in the past decade.¹⁰

Legal commentators have responded to the growing research about the pervasiveness of implicit bias and the poor success record of plaintiffs with proposals that would expand liability standards and make it easier for plaintiffs to prove discrimination in Title VII cases. Some scholars favor use of presumptions that would require findings of discrimination, even when explanations other than discrimination are no less likely. One scholar suggests a new liability category for ambiguous circumstances that “outsiders” would recognize as discriminatory although “insiders” would not. Others would impose liability on employers for the failure to adopt workplace procedures and structures that *might* have prevented discrimination from occurring, even when the link between existing procedures and structures and actual race or gender bias cannot be proved.¹¹

making, and Misremembering, 57 Duke L.J. 345 (2007) (suggesting that judges and jurors misremember case facts in racially biased ways).

⁹ See Kevin M. Clermont & Stewart J. Schwab, Employment Discrimination Plaintiffs in Federal Court: From Bad to Worse?, 3 Harv. L. & Pol’y Rev. 103, 110 display 2 (2009) (defendants are almost three times more likely to win on appeal of plaintiffs’ wins at pretrial than plaintiffs are to win on appeal of defendants’ pretrial wins, and defendants are more than four times more likely than plaintiffs to win appeals of trial verdicts); Kevin M. Clermont & Stewart J. Schwab, How Employment Discrimination Plaintiffs Fare in Federal Court, 1 J. Empirical Legal Studies 429 (2004) (employment discrimination plaintiffs are four times as likely to lose in the pre-trial litigation stage as plaintiffs in other civil cases, twice as likely to lose at trial, and four times as likely to have their verdicts overturned on appeal). The win rate of Title VII cases filed in federal court between 1998 and 2006 was just over ten percent. Clermont & Schwab, From Bad to Worse?, supra, at 117 display 6.

¹⁰ See Clermont & Schwab, From Bad to Worse?, supra note 9, at 115–18 (termination of federal employment discrimination cases fell from 23,721 in 1999 to 15,007 in 2007); Laura Beth Nielsen et al., Uncertain Justice: Litigating Claims of Employment Discrimination in the Contemporary United States 14 (ABA Research Paper No. 08-04, Apr. 16, 2008), available at <http://ssrn.com/abstract=1093313> (federal employment discrimination filings dropped from 23,796 in fiscal year 1997 to 14,353 in fiscal year 2006).

¹¹ See infra Section II.B.

That more aggressive legal solutions to address the elusive problem of implicit discrimination are sought is not surprising. Title VII has significantly reduced workplace discrimination; perhaps if the law can be made even stricter, it can eliminate even more of it. More of a good thing, however, is not always better. One problem with relying on more law to reduce discrimination, already noted by other scholars, is that law is an ineffective instrument for eliminating behaviors we cannot readily define or correct. Amy Wax argues that imposing liability on employers for a phenomenon they cannot discern and do not know how to prevent leads them to overinvest in precautions that may reduce the danger of liability while not actually reducing the influence of unconscious bias on decisionmaking.¹² Another problem, about which Susan Sturm has written extensively, is that individual lawsuits address only individual behaviors and not the structural features of the workplace that enable discrimination to occur. According to Sturm, broader institutional strategies are necessary to confront the structural forces that mask and facilitate discrimination in the workplace and channel decisionmaking into more nondiscriminatory patterns.¹³ Still another response, made forcefully by Ralph Richard Banks and Richard Thompson Ford, is that the focus on implicit bias is misplaced because the law already treats unconscious and conscious bias the same and because unconscious bias discourse confuses and derails the real goal of anti-racism, which is the elimination of entrenched racial inequalities.¹⁴

¹² See Amy L. Wax, *Discrimination as Accident*, 74 *Ind. L.J.* 1129, 1133 (1999). Wax argues that Title VII covers unconscious as well as conscious discrimination, but still requires that discrimination be proved. See Amy L. Wax, *The Discriminating Mind: Define It, Prove It*, 40 *Conn. L. Rev.* 979 (2008).

¹³ See, e.g., Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 *Colum. L. Rev.* 458, 460–61 (2001); see also Susan Sturm, *The Architecture of Inclusion: Advancing Workplace Equity in Higher Education*, 29 *Harv. J.L. & Gender* 247 (2006) (suggesting that institutional reform is more promising than legal reform in making universities more inclusive).

¹⁴ Ralph Richard Banks & Richard Thompson Ford, (How) Does Unconscious Bias Matter?: Law, Politics, and Racial Inequality, 58 *Emory L.J.* 1053, 1058–59, 1072–89, 1113–21 (2009); see also Charles Lawrence III, *Unconscious Racism Revisited: Reflections on the Impact and Origins of “The Id, the Ego, and Equal Protection,”* 40 *Conn. L. Rev.* 931, 942 (2008) (arguing that the emphasis on unconscious bias has diverted attention from the ideology and material structures of white supremacy).

This Article assumes that unconscious discrimination is a problem worthy of focus, both from the perspective of achieving non-discrimination in the workplace and in reducing the larger societal disparities based on race and gender. It also sides with those scholars who believe that Title VII already prohibits unconscious as well as conscious race and gender discrimination.¹⁵ The Article's purpose, however, is not to establish the scale and significance of the unconscious discrimination problem, or to defend existing law as ideal. Instead, it cautions against approaches to unconscious discrimination—whatever its prevalence and whatever the inadequacies of existing law—that rely principally on stronger legal coercion as the primary tool to fight implicit discrimination. The problem the Article identifies is that coercion, despite its necessity in many circumstances, can have a negative effect on people's internalization of nondiscrimination norms. Aggressive legal strategies assume that the stronger the legal pressure, the more effective it will be in reducing undesirable behaviors. This Article makes the case that, specifically in the context of behaviors based on implicit bias, just the opposite may be true.

The potentially counter-productive nature of law with respect to implicit bias is rooted in many factors, not the least of which is that people are not aware of their biases, and are highly sensitive to circumstances in which others may think that they are prejudiced. When caught (by themselves or by others) doing or saying something that might be viewed as racist or sexist, people tend to respond by saying (or thinking to themselves), "I didn't mean it that way." How others respond to this reaction itself reflects a profound difference in perspective about nondiscrimination goals and the way to achieve them. Some see such protestations of innocence as a form of self-deception or self-rationalization that should be exposed.¹⁶ People are prone to have self-serving, negative thoughts

¹⁵ See Banks & Ford, *supra* note 14, at 1072–89 (arguing that Title VII may not be adequate, but the problem is not that it treats conscious and unconscious bias differently); Amy Wax, *Discrimination as Accident*, 74 *Ind. L.J.* 1129, 1146–52 (1999) (reading Title VII and the cases applying it to prohibit both conscious and unconscious discrimination).

¹⁶ See Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 *Stan. L. Rev.* 317, 387 (1987) (implying that the intent requirement in antidiscrimination law rationalizes the status of the privileged who do not intentionally discriminate); David Wellman, *Unconscious Racism, Social Cogni-*

2009]

Making Good on Good Intentions

1901

that will find expression in their actions if not apprehended and corrected. Only when people are confronted and held liable for their prejudice, whenever and however it can be identified, will they stop discriminating. The important thing, this line of reasoning goes, is to name the wrong and hold people accountable for it. Anything less simply enables bias to flourish. Since discrimination is ubiquitous, there is little chance we will overcorrect for it.¹⁷

An alternative theory—the one elevated by this Article—is that threat and confrontation about race and gender bias, which people do not want to possess or exhibit, may inadvertently provoke shame, guilt, and resentment, which lead to avoidance and resistance, and ultimately to more stereotyping. In other words, pressure and threat will often deepen bias rather than correct it. Positive strategies that affirm people's good intentions, in contrast, engage people constructively in defining their better, nondiscriminatory selves and aligning their conduct accordingly. While coercion and threat make people defensive, opportunity and engagement leverage people's good intentions into a deeper commitment to a more inclusive, nondiscriminatory workplace. It is this type of commitment—not legal coercion—that will best address the implicit bias that is most characteristic of today's workplace.

Much scholarly attention has already been given to the need for strong legal standards to combat today's workplace discrimination.¹⁸ This Article endorses the need for strong standards, but argues that attention also needs to be given to the means by which internal commitment to those standards, or what I refer to as *good intentions*, is generated. Some researchers have suggested that good intentions, even if effective against explicit forms of bias, are

tion Theory, and the Legal Intent Doctrine: The Neuron Fires Next Time, *in* Handbook of the Sociology of Racial and Ethnic Relations 39 (Hernán Vera & Joe R. Feagin eds., 2005) (exposing the extent to which racial advantage is invisible and rationalized); see also Banks & Ford, *supra* note 14, at 1104–05 (criticizing unconscious bias discourse because it permits people to affirm their self-image as non-racist).

¹⁷ See Ann C. McGinley, ¡Viva La Evolucion!: Recognizing Unconscious Motive in Title VII, 9 Cornell J.L. & Pub. Pol'y 415, 482 (2000).

¹⁸ Among some recent examples, see Tristin K. Green, A Structural Approach as Antidiscrimination Mandate: Locating Employer Wrong, 60 Vand. L. Rev. 849, 851–53 (2007); Russell K. Robinson, Perceptual Segregation, 108 Colum. L. Rev. 1093, 1103 (2008).

largely ineffective to stop implicit discrimination.¹⁹ This Article posits, to the contrary, that good intentions have their greatest comparative advantage when it comes to the more subtle forms of discriminatory behavior, and that people who have an internal commitment to nondiscrimination norms will combat implicit discrimination more effectively than those motivated by traditional legal sanctions. The Article also points toward strategies that will best encourage that internal commitment.

Social science research suggests that people are most likely to internalize norms when they feel autonomous, competent, and related to others.²⁰ Coercion threatens these core prerequisites of the internalization process because it leads people to feel controlled, untrusted, and alienated. Coercion with respect to race and gender bias seems to be particularly threatening. Thus, while clear, firm, and enforceable legal standards are necessary in order to define basic limits on discriminatory behavior, when these standards come to feel unfair or overly controlling, they evoke guilt, resentment, and resistance—all reactions that actually increase stereotyping. When people feel that the rules are unfair or coercive, the most that usually can be achieved is rote compliance. Rote compliance by those who are resigned or indifferent, perhaps even hostile, to nondiscrimination norms may be enough with respect to explicit forms of discrimination, but it will not achieve significant headway against implicit discrimination. To reach beyond such behaviors requires people who are committed to nondiscrimination norms and determined to live by them.

¹⁹ See, e.g., John A. Bargh, *The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects*, in *Dual-Process Theories of Social Psychology* 361, 361, 362, 376, 378 (Shelly Chaiken & Yaacov Trope eds., 1999) (noting there is little that can be done to control influence of automatic stereotypes; the deactivation of unconscious stereotypes would require the almost unimaginable combination of a person being aware of the nonconscious influences, having a relatively accurate theory about the nature of those influences, and being sufficiently motivated and skilled to correct for them); Kang & Banaji, *supra* note 2, at 1079 (“[A]n explicit ex ante exhortation not to be intentionally unfair will do little to counter implicit cognitive processes, which take place outside our awareness yet influence our behavior.”).

²⁰ See Edward L. Deci & Richard M. Ryan, *The “What” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior*, 11 *Psychol. Inquiry* 227 (2000) (citing studies); see *infra* Subsection III.B.2.

2009]

Making Good on Good Intentions

1903

In focusing on good intentions, the goal of the Article is not to argue that people who engage in unintended discrimination should be let off the hook. Title VII, correctly read, prohibits discrimination based on race, sex, and other protected characteristics, whether or not the discrimination was intended. Where cases of implicit discrimination can be proved, defendants should be held responsible. The Article recognizes that many cases of implicit discrimination cannot be readily proved. It argues, however, that this fact has more to do with the complexities of implicit discrimination than with deficiencies in existing law. Too much of the current academic discussion assumes from the prevalence of implicit bias that tougher laws are necessary, and would be able to, combat it. Looking more closely at the social science about the effects of coercion on reducing discrimination, this Article demonstrates that the matter is not so simple.

The Article may appear to focus on changing potential discriminators at the expense of justice for deserving victims. But it is unrealistic to think that making it substantially easier to prove discrimination will either promote justice or reduce the kinds of discrimination that people do not intend and of which they are not aware. Even if some benefits might be expected, any short-term gains must be weighed against the long-term effects of more legal pressure, which requires attention to the impact of these rules on people's attitudes and beliefs. Without proposing to make anything worse for plaintiffs or weakening Title VII in its present form, the Article outlines the principles that might help us to think about that impact.

The Article may also seem naïvely optimistic about the changeability of human nature. If a change of attitudes is to be a possibility, however, that possibility must be entertained. Good intentions are not the only factor that matters in reducing discrimination,²¹ and the Article makes clear that they should not be a free pass to discrimination,²² but they are a form of social capital that should be fostered, like any other asset. They can be nurtured, or they can be squandered. It pays to pay attention to the difference.

²¹ Nor are good intentions sufficient. They may, however, make the difference between whether or not other specific strategies to reduce bias actually work. See *infra* Part IV.

²² See *infra* Section III.A.

This Article will proceed as follows. Part I will examine the psychological dimensions of workplace discrimination. It will describe both the cognitive and the motivated processes of dimensions of bias and stereotyping, as they contribute to the general phenomenon of implicit discrimination. It will demonstrate that, despite the near-exclusive emphasis on cognitive stereotyping in the legal literature, motivation is a key component in the processes entailed in implicit discrimination. Part II will address the law of employment discrimination, briefly explaining the extent to which Title VII already prohibits implicit as well as explicit discrimination. This Part will also describe proposed reforms to existing law that are aimed at making Title VII more effective in combating unintended discrimination. Part III will review social science literature about the role of motivation in reducing stereotyping. This Part will stress (1) the significance of the difference between external and internal motivation to norm compliance, (2) the importance of autonomy, competence and connectedness to a person's ability to internalize norms, and (3) the positive and the negative roles the law can play in enhancing people's intentional commitment to comply with nondiscrimination norms. Part IV will examine the social science research evaluating measures to reduce prejudice. Some of the research focuses on the role of individual effort and attention in reducing bias, and some of it focuses on situational influences ranging from counterstereotype images to peer interaction, social norms, and workplace intergroup contact. The research demonstrates that the effects of both individual effort and situational change depend significantly on whether people are externally or internally motivated to reduce prejudice. Part V will analyze the implications of these research findings for legal and workplace strategies to reduce implicit discrimination.

I. THE PSYCHOLOGICAL DIMENSIONS OF RACE AND GENDER DISCRIMINATION

There is general agreement that discrimination based on explicitly negative attitudes about race and sex has diminished over time,²³ and that a growing portion of discrimination is unconscious

²³ Sources documenting the decline in racism and sexism since the 1940s include Richard J. Crisp & Rhiannon N. Turner, *Essential Social Psychology* 166–68 (2007);

2009]

Making Good on Good Intentions

1905

and unintended.²⁴ Some believe that discrimination changes form more than it actually diminishes.²⁵ This belief is often associated with an account of discrimination as an ideological instrument of

Howard Schuman et al., *Racial Attitudes in America: Trends and Interpretations* 104–05 (1997); John F. Dovidio, *On the Nature of Contemporary Prejudice: The Third Wave*, 57 *J. Soc. Issues* 829, 930–31 (2001); Peter Glick & Susan T. Fiske, *The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism*, 70 *J. Personality & Soc. Psychol.* 491 (1996).

As more explicit forms of racism and sexism declined, so evolved the terms used to name each. Thus, what was once considered “racism” later came to be referred to as “old-fashioned racism” or “dominative racism,” and was then replaced by terms encompassing more subtle forms, called “modern racism,” and later “symbolic,” “ambivalence,” “aversive,” and (more generally) “implicit” racism. See Joel Kovel, *White Racism: A Psychohistory* 54 (1970) (discussing dominative racism); John Duckitt, *The Social Psychology of Prejudice* 19–24 (1992) (discussing symbolic racism); Samuel L. Gaertner et al., *Aversive Racism: Bias Without Intention*, in *Handbook of Employment Discrimination Research* 377 (Laura Beth Nielsen & Robert L. Nelson eds., 2005) (discussing aversive racism); see also John F. Dovidio, *On the Nature of Contemporary Prejudice: The Third Wave*, 57 *Soc. Issues* 829 (2001) (giving a historical overview of study of racism, with focus on aversive racism).

The study of sexism over time has produced a similar progression of terms. See Janet K. Swim et al., *Sexism and Racism: Old-Fashioned and Modern Prejudices*, 68 *J. Personality & Soc. Psychol.* 199 (1995) (comparing Old-Fashioned Sexism and Modern Sexism); Francine Tougas et al., *Neo-Sexism: Plus Ça Change, Plus C’est Pareil*, 21 *Personality & Soc. Psychol. Bull.* 842 (1995) (discussing neo-sexism); Peter Glick et al., *Beyond Prejudice as Simple Antipathy: Hostile and Benevolent Sexism Across Cultures*, 79 *J. Personality & Soc. Psychol.* 763 (2000) (comparing hostile and benevolent sexism); Glick & Fiske, *supra* (discussing ambivalent sexism).

²⁴ Several scholars have observed that the trends in today’s workplace present more opportunities for the subtle forms of discrimination than were possible in prior generations. These trends include the flattening of workplace hierarchies, blurring of job boundaries, allocation of work in teams rather than to individuals, adoption of more skill-based, individualistic, and flexible methods of workplace evaluation, and reduction in job security. See Katherine V. W. Stone, *From Widgets to Digits: Employment Regulation for the Changing Workplace* 174–83 (2004); Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 *Harv. C.R.-C.L. L. Rev.* 91, 93, 101–03 (2003); see also Cynthia Estlund, *Working Together: How Workplace Bonds Strengthen a Diverse Democracy* (2003).

²⁵ See, e.g., Faye Crosby et al., *Recent Unobtrusive Studies of Black and White Discrimination and Prejudice: A Literature Review*, 87 *Psychol. Bull.* 546 (1980) (documenting continued racial bias, in the face of public opinion polls appearing to show the decline of racism); Olatunde Johnson, *Disparity Rules*, 107 *Colum. L. Rev.* 374 (2007) (arguing that continued racial disparities show that racism still exists); Lincoln Quillian, *New Approaches to Understanding Racial Prejudice and Discrimination*, 32 *Ann. Rev. Soc.* 299, 302–09 (2006) (describing audit studies that show a persistence of race discrimination in housing and employment, despite evidence that negative racial attitudes have declined).

1906

Virginia Law Review

[Vol. 95:1893]

group power,²⁶ in which privileged groups take “strategic, self-interested actions” to “intentionally exclude and exploit subordinate-group members [in order] to protect or advance their own interests.”²⁷ By this account, the privileged rationalize their subordination of others through self-serving ideologies that change with the time. Under Jim Crow, segregation was justified on the basis of the inferiority of blacks,²⁸ and the failure to give women the vote was rationalized by the separate spheres of men and women dictated by God and nature.²⁹ Today’s racist and sexist practices are said to be sustained more subtly, but no less insidiously, through purportedly neutral principles such as equality, difference, and the free market.³⁰ Even the victims of discrimination are drawn to these ideologies, thereby helping to perpetuate their own subordination. According to system justification theory, believing that one lives in

²⁶ See Lawrence D. Bobo, Prejudice as Group Position: Microfoundations of a Sociological Approach to Racism and Race Relations, 55 *J. Soc. Issues* 445, 447 (1999) (developing a group position approach); Barbara F. Reskin, The Proximate Causes of Employment Discrimination, 29 *Contemporary Soc.* 319, 321 (2000) (concluding that sociologists see prejudice as a “fundamental mechanism of stratification through which dominant groups preserve their privileged position”).

²⁷ Reskin, *supra* note 26, at 320.

²⁸ See Bobo, *supra* note 26, at 464 (describing the Jim Crow belief that blacks are inferior to whites).

²⁹ See *Bradwell v. Illinois*, 83 U.S. (16 Wall.) 130, 141 (1872) (Bradley, J., concurring) (separate spheres of men and women is “founded in the divine ordinance, as well as in the nature of things”).

³⁰ In the context of race, Lawrence Bobo refers to this as “laissez-faire racism.” See Bobo, *supra* note 26, at 464–65. In the context of sex, Catharine A. MacKinnon calls the theory that neutral principles such as objectivity, equality, and free speech are constructed so as to protect men’s interests and subordinate women “feminism unmodified.” See Catharine A. MacKinnon, *Feminism Unmodified: Discourses on Life and Law* 48–62 (1987); Catharine A. MacKinnon, *Toward a Feminist Theory of the State* 157–70 (1989). There are a number of variations on this theme. See, e.g., David O. Sears et al., *Egalitarian Values and Contemporary Racial Politics*, in *Racialized Politics: The Debate About Racism in America* 75, 79 (David O. Sears et al. eds., 2000) (suggesting that negative views toward race today are fused with race-neutral traditional American values, such as individualism, the work ethic, delay of gratification, patriotism, and respect for authority); Jim Sidanius & Felicia Pratto, *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression* (1999) (tying opposition to race-targeted policies to the desire for group dominance); Paul M. Sniderman & Edward G. Carmines, *Reaching Beyond Race*, 30 *PS: Political Science & Politics* 466, 471 (1997) (stating that differences in racial policy beliefs today are about politics, not race).

2009]

Making Good on Good Intentions

1907

a rational, fair, and orderly world is often preferable to believing that one is a victim in a fundamentally unjust society.³¹

This group-position account of discrimination helps explain the sustenance of power and privilege at a deep institutional level, even in societies that appear to be moving in a politically progressive direction. On its own terms, however, this account limits potential strategies for reducing discrimination to the pursuit of convergent interests between those who have power and those who do not,³² or to punishing those who violate the favorable rules that emerge from these political deals. For the most part, it takes people at their self-interested worst, and suggests no mechanisms or strategies for transformative attitudinal and institutional change

Acknowledging the force of this theory, this Article is premised on an alternative, more optimistic account, which relies on more

³¹ See Gary Blasi & John T. Jost, System Justification Theory and Research: Implications for Law, Legal Advocacy, and Social Justice, 94 Cal. L. Rev. 1119 (2006); John T. Jost & Mahzarin R. Banaji, The Role of Stereotyping in System Justification and the Production of False Consciousness, 33 Brit. J. Soc. Psychol. 1 (1994). System justification theory, although rooted in social psychology rather than sociology, is congruent with other ideology-based theories according to which people reconcile themselves to the world they live in, rather than confront its injustices, such as the theory of cognitive dissonance, and the “just world” thesis. See, e.g., Leon Festinger, *A Theory of Cognitive Dissonance* (1957); Melvin J. Lerner, *The Belief in a Just World: A Fundamental Delusion* (1980). Belief in a just world lessens the likelihood that members of lower-status groups will perceive discrimination when rejected for a job position, and increases the chances that members of higher-status groups will blame their rejection on discrimination. See Deborah L. Brake, Perceiving Subtle Sexism: Mapping the Social-Psychological Forces and Legal Narratives that Obscure Gender Bias, 16 Colum. J. Gender & L. 679, 688 (2007) (surveying research); see also MacKinnon, *Toward a Feminist Theory of the State*, supra note 30, at 114 (arguing that women are pushed by the dominant, male perspective to see reality in male terms that denies their oppression, even though it contradicts some of their lived experience); Peter Glick et al., Beyond Prejudice as Simply Antipathy: Hostile and Benevolent Sexism Across Cultures, 79 J. Personality & Soc. Psychol. 763, 773 (2000) (women adopt stereotypes about themselves as a form of self-defense and survival). For discussion of how ideologies and situational factors are “mutually reinforcing” in stimulating group competition and prejudice, see Victoria M. Esses et al., Instrumental Relations Among Groups: Group Competition, Conflict, and Prejudice, in *On the Nature of Prejudice: Fifty Years After Allport* 227, 233 (John F. Dovidio et al. eds., 2005).

³² The interest-convergence thesis is set forth in Derrick Bell, *Diversity’s Distractions*, 103 Colum. L. Rev. 1622, 1624 (2003); see also Derrick A. Bell Jr., *Brown v. Board of Education and the Interest-Convergence Dilemma*, 93 Harv. L. Rev. 518, 523 (1980) (suggesting that progress of blacks occurs only when it serves the interests of whites). For a critique, see Justin Driver, *Rethinking the Interest-Convergence Thesis* (unpublished manuscript, on file with the Virginia Law Review).

nuanced and less totalizing explanations for race and gender discrimination. This account allows for the existence of well-intended people, assumes they can make a difference, treats people's attitudes and behaviors as changeable, and posits that social and institutional contexts can influence that change. It thus emphasizes how best to bring about this change.

This Part reviews basic social science about stereotyping, prejudice, and discrimination. It uses the standard social science distinctions between a *stereotype*, which is a faulty generalization about a group or its members; *prejudice* or *bias*, which is the positive or negative attitude that can attach to a stereotype; and *discrimination*, which is negative treatment based on stereotype and prejudice that may follow.³³ In studying these phenomena, some scholars have distinguished single acts of biased individuals from the larger and more significant institutional context in which discrimination is enabled.³⁴ This Article treats the relationship between the two as the critical factor. The attitudes and beliefs of individuals about race and gender not only are shaped by, but also shape, the social and institutional contexts of which they are part. Neither can be improved without some understanding of the interaction with the other. This Part focuses primarily on the processes of individual stereotyping. Parts III and IV will put individual stereotyping in broader situational context.

A. Stereotyping as Social Cognition

Stereotypes are categories that constrain and shape what a person believes about, and expects from, other people.³⁵ According to social cognition theory, people are drawn to those categories that are the most visible, or salient, such as gender, race, and age.³⁶ Con-

³³ Robert L. Dipboye & Adrienne Colella, An Introduction, in *Discrimination at Work: The Psychological and Organization Bases 1-2* (Robert L. Dipboye & Adrienne Colella eds., 2005).

³⁴ See, e.g., Tristin K. Green & Alexandra Kalev, *Discrimination-Reducing Measures at the Relational Level*, 59 *Hastings L.J.* 1435, 1450-52 (2008); Sturm, *Second-Generation Discrimination*, supra note 13, at 466-71; Wellman, supra note 16, at 53.

³⁵ Diane M. Mackie et al., *Social Psychological Foundations of Stereotype Formation*, in *Stereotypes and Stereotyping* 41, 42-43 (C. Neil Macrae et al. eds., 1996).

³⁶ See Susan T. Fiske et al., *The Continuum Model: Ten Years Later*, in *Dual-Process Theories*, supra note 19, at 231, 232-34 (stating that gender, ethnicity and age are "privileged" social categories because they are immediately physically manifested

2009]

Making Good on Good Intentions

1909

text can increase or decrease category salience. For example, a black female's gender is likely to be salient in a group of men, whereas her race will tend to be more salient in a group of white women.³⁷

Categories help to sort, store, and retrieve information, with an efficiency that would not be possible if each person, object, and situation had to be encountered as an entirely individualized phenomenon.³⁸ Useful as they are, however, categories also distort

and have important cultural meanings that are often relevant to people's immediate interaction goals); David L. Hamilton et al., *Social Cognition and the Study of Stereotyping*, in *Social Cognition: Impact on Social Psychology* 291, 311–12 (Patricia G. Devine et al. eds., 1994) (citing research suggesting that these categories are based on basic and unalterable biological differences and reflect other differences on a variety of other attributes). The importance of these traits as organizational categories is apparent even in young children. See, e.g., Andrew Scott Baron & Mahzarin R. Banaji, *The Development of Implicit Attitudes: Evidence of Race Evaluations from Ages 6 and 10 and Adulthood*, 17 *Psychol. Sci.* 53, 56 (2006). Children appear to use gender even earlier than race or age to categorize people. See Mackie et al., *supra* note 35, at 46–47.

The commonality of gender, race, and age as cross-cultural categories does not mean that perceptual tendencies are uniform across cultures. For example, some research indicates that East Asians perceive objects and people more situationally, in their contextual frameworks, than Americans, while Americans are more likely to make assumptions about people's dispositions and personality traits. For a review of some of the studies and some original research involving Japanese research subjects, see Takahiko Masuda & Richard E. Nisbett, *Attending Holistically Versus Analytically: Comparing the Context Sensitivity of Japanese and Americans*, 81 *J. Personality & Soc. Psychol.* 922 (2001).

³⁷ See Jason P. Mitchell et al., *Contextual Variations in Implicit Evaluation*, 132 *J. Experimental Psychol.: Gen.* 455, 459 (2003). The more extreme the characteristics, the more salient, and thus the more likely they are to trigger stereotypes. See Irene V. Blair et al., *The Role of Afrocentric Features in Person Perception: Judging by Features and Categories*, 83 *J. Personality & Soc. Psychol.* 5 (2002) (finding that skin tone, eyes, nose shape and size, lips, and hair texture determined how quickly blacks are categorized by race, and thus to be judged by racial stereotypes); Devon Carbado & Mitu Gulati, *Working Identity*, 85 *Cornell L. Rev.* 1259 (2000) (suggesting that race-associated behaviors enhance stereotyping).

³⁸ As stated by Gordon Allport, who is widely credited with the development of social categorization theory upon which others have built, "The human mind must think with the aid of categories [or] *generalizations*. Once formed, categories are the basis for normal prejudice. We cannot possibly avoid this process. Orderly living depends upon it." Gordon Allport, *The Nature of Prejudice* 20 (1954). Stereotypes also serve social functions such as by signaling whether one should offer one's seat on a crowded bus (to, say, a disabled or elderly person), or whether it is all right to ask someone for a glass of water (say, a waitress), a book (a librarian), or for help (a police officer). See Mark Snyder & Peter Miene, *On the Functions of Stereotypes and*

people's judgments about the world. For one thing, they cause people to perceive things and people within a category as being more alike than if the category did not exist, and those in different categories as more different.³⁹ This tendency has been demonstrated in experiments in which people estimate lines grouped in the same category as more similar in length, and more different from lines in other groups, than they actually are.⁴⁰ They do the same thing with people.⁴¹

Categories also lead people to attribute to all members of a category characteristics possessed by only some of them. In other words, people over-generalize from a "representative" few to the many.⁴² As put by Gordon Allport in 1954, a child thinks that "all houses must have, as his house has, two floors, a gas refrigerator, and a television set."⁴³ So, too, people often assume that other people who are alike in one respect, like race or sex, are also alike in others.⁴⁴

Once generalizations based on categories are fixed in people's minds, the way they process information perpetuates those generalizations. They notice, ask questions about, remember, and integrate information that confirms their previously formed views,⁴⁵

Prejudice, *in* 7 *The Psychology of Prejudice: The Ontario Symposium*, 33, 46 (Mark P. Zanna & James M. Olson eds., 1994).

³⁹ See Samuel L. Gaertner & John F. Dovidio, *Reducing Intergroup Bias: The Common Ingroup Identity Model* 35 (2000). Applied to social groups, this is called the "out-group homogeneity effect." See Thomas M. Ostrom & Constantine Sedikides, *Out-Group Homogeneity Effects in Natural and Minimal Groups*, 112 *Psychol. Bull.* 536 (1992).

⁴⁰ See Henri Tajfel, *Cognitive Aspects of Prejudice*, 25 *J. Soc. Issues*, Autumn 1969, at 79, 83-86 (1969); Henri Tajfel & A.L. Wilkes, *Classification and Quantitative Judgement*, 54 *Brit. J. Psychol.* 101, 104 (1963).

⁴¹ See *infra* text accompanying notes 56-58.

⁴² These overgeneralized attributions are called illusory correlations. See David L. Hamilton et al., *The Influence of Affect on Stereotyping: The Case of Illusory Correlations*, in *Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception* 39, 44 (Diane M. Mackie & David L. Hamilton eds., 1993).

⁴³ Allport, *supra* note 38, at 171. Allport reports that "[f]or many years Americans *imagined* that all Bolsheviks wore whiskers." *Id.* at 132.

⁴⁴ Citing some of the basic sources, see Susan T. Fiske, *Stereotyping, Prejudice, and Discrimination*, *in* 2 *The Handbook of Social Psychology* 357, 362 (Daniel T. Gilbert et al. eds., 1998).

⁴⁵ For studies of "confirmation bias," see Steven L. Neuberg, *Expectancy-Confirmation Processes in Stereotype-Tinged Social Encounters: The Moderating Role of Social Goals*, *in* 7 *The Psychology of Prejudice: The Ontario Symposium* 103,

2009]

Making Good on Good Intentions

1911

while they tend to ignore, forget, or explain away contradictory information.⁴⁶

B. Stereotyping as Motivated Prejudice

In a widely influential article, Linda Hamilton Krieger argues that intergroup bias is largely a result of cognitive functioning, of the type just summarized.⁴⁷ As such, she argues, it is unintentional, and thus not captured by Title VII, which she claims does not cover unintended discrimination.⁴⁸ I address later what I believe to be the overstated claim that Title VII covers only intended discrimination.⁴⁹ In this Section, I focus on the social science research demonstrating that categorization is not only a cognitive mechanism, but also a motivated one. This point is important because if motivation is an integral part of the stereotyping process, strategies to reduce stereotyping and discrimination must address the motivational components as well as the cognitive ones.

People bring various motivations to how they view others, especially the need to belong, to control, to possess, to understand, to feel safe, and to feel good about themselves.⁵⁰ Whether these needs are met affects people's responses to others, including whether they judge others based on stereotypes.

A principal form of motivated bias is the tendency for people to classify themselves into one social category and out of others, and

108 (1994). Because of this bias, it is not necessary for a thing or person to *be* different; it is enough that they *appear to be* different. See Allport, *supra* note 38, at 132.

⁴⁶ See Fiske, Stereotyping, Prejudice, and Discrimination, *supra* note 44, at 368. Researchers have referred to the process by which the impact of disconfirming information may be diffused as "prototype subtyping." See Miles Hewstone, Contact and Categorization: Social Psychological Interventions to Change Intergroup Relations, *in* *Stereotypes and Stereotyping* 323, 338–41 (C. Neil Macrae et al. eds., 1996); see also sources in *infra* note 55.

⁴⁷ See Krieger, *supra* note 7, at 1165.

⁴⁸ *Id.* at 1239.

⁴⁹ See *infra* Section II.A.

⁵⁰ See Susan T. Fiske, Intent and Ordinary Bias: Unintended Thought and Social Motivation Create Casual Prejudice, 17 *Soc. Just. Res.* 117, 123–24 (2004). For a fuller discussion of human needs that bear on prejudice, see Section III.A. For sources describing earlier understandings of the needs that produced race prejudice as pathologies, see John Duckitt, *The Social Psychology of Prejudice* 52–54 (1992); Fiske, *supra* note 44, at 358–59.

to favor those in their own category.⁵¹ This tendency is often referred to as “ingroup” bias, which is mirrored by the corresponding tendency to disfavor members of the “outgroup.”⁵² People bring their biases favoring members of their ingroup and disfavoring outgroup members to their cognitive tendencies to make illusory correlations, confirm existing biases, and view things in the same category as more alike than they actually are.⁵³ These processes affect how people notice, digest, and remember information about others.⁵⁴ For example, people tend to notice and retain more detailed information for ingroup members; in interview settings they ask more questions and the questions are designed to obtain individuating information about ingroup members. In contrast, people tend to notice and retain information about outgroup members that conforms to stereotypes about that group; in interview settings, they ask fewer questions and those they ask are designed to confirm what they expect, not change it.⁵⁵ As Steven Neuberg states, when people hold strong stereotypes of certain groups, “they often cannot help but take note of the many individuals who are indeed as they expected them to be.”⁵⁶

People exhibit ingroup bias by responding to members of their own ingroup with greater empathy, respect, and cooperation. In experimental studies, people distribute more resources to ingroup

⁵¹ Samuel L. Gaertner & John F. Dovidio, *Reducing Intergroup Bias: The Common Ingroup Identity Model* 36 (2000).

⁵² Marilyn B. Brewer, *The Psychology of Prejudice: Ingroup Love or Outgroup Hate?*, 55 *J. Soc. Issues* 429, 430 (1999). Researchers have demonstrated that while ingroup and outgroup biases are the converse of each other, ingroup bias may exist even in the absence of negative attitudes toward outgroups. *Id.* at 432.

⁵³ See *supra* text accompanying notes 39–41.

⁵⁴ See Mackie et al., *supra* note 35, at 46 (citing examples from a number of studies); Fiske, *Stereotyping, Prejudice, and Discrimination*, *supra* note 44, at 368–69 (citing studies).

⁵⁵ See Steven L. Neuberg, *The Goal of Forming Accurate Impressions During Social Interactions: Attenuating the Impact of Negative Expectancies*, 56 *J. Personality & Soc. Psychol.* 374, 375 (1989); see also Mark Snyder & William B. Swann, Jr., *Hypothesis-Testing Processes in Social Interaction*, 36 *J. Personality & Soc. Psychol.* 1202 (1978) (describing a study showing how people use social interactions to confirm their hypotheses about other people and how those social interactions also serve to bring about the expected behaviors).

⁵⁶ Neuberg, *Expectancy-Confirmation Processes in Stereotype-Tinged Social Encounters: The Moderating Role of Social Goals*, *supra* note 45, at 106.

2009]

Making Good on Good Intentions

1913

members, and they evaluate them more positively.⁵⁷ They also are inclined to be more generous in their explanations for the behavior of ingroup members than they are towards others. For example, people tend to attribute an achievement of someone in their own group to the person's abilities or dispositions, while they attribute an achievement by others to luck or some other external factor.⁵⁸ Conversely, although to a somewhat lesser extent, they view negative behaviors and outcomes as flukes or exceptions when exhibited by ingroup members, and as internal, stable characteristics when evidenced in outgroup members.⁵⁹ People also tend to judge behaviors as more negative and intentional when performed by someone outside their group; an action that is perceived as a simple push by a member of one's own group may be perceived as an act of aggression when performed by someone else.⁶⁰

Some researchers hypothesize that people favor members of their own ingroups as part of a larger preference for the familiar, which is more comfortable, less challenging, and less threatening than the unfamiliar.⁶¹ This hypothesis is consistent with Krieger's

⁵⁷ See research collected in Gaertner & Dovidio, *Reducing Intergroup Bias: The Common Ingroup Identity Model*, supra note 51, at 38–39.

⁵⁸ See Kay Deaux & Marianne LaFrance, *Gender*, in 1 *The Handbook of Social Psychology* 788, 798 (Daniel T. Gilbert et al. eds., 4th ed. 1998) (women's success at traditionally male tasks tends to be attributed to effort, whereas men's comparable success is attributed to ability). This phenomenon has been labeled the "ultimate attribution error." See Miles Hewstone, *The "Ultimate Attribution Error": A Review of the Literature on Intergroup Attributions*, 20 *Eur. J. Soc. Psychol.* 311 (1990) (reviewing nineteen studies documenting attribution errors, especially in the context of racial groups); Thomas F. Pettigrew, *The Ultimate Attribution Error: Extending Allport's Cognitive Analysis of Prejudice*, 5 *Personality & Soc. Psychol. Bull.* 461, 464–65 (1979) (defining phenomenon).

⁵⁹ See Hewstone, *The "Ultimate Attribution Error": A Review of the Literature on Intergroup Attributions*, supra note 58.

⁶⁰ Negative action by an out-group member is also perceived at a more abstract and stable level ("she is hostile") than identical behavior of the ingroup member ("she slapped the girl"), while the reverse is true about positive behaviors ("she walked across the street holding the man's hand" versus "she is helpful"). See Gaertner & Dovidio, *Reducing Intergroup Bias: The Common Ingroup Identity Model*, supra note 51, at 39; Hewstone, *The "Ultimate Attribution Error": A Review of the Literature on Intergroup Attributions*, supra note 58; Anne Maass et al., *Linguistic Intergroup Bias: Evidence for In-Group-Protective Motivation*, 71 *J. Personality & Soc. Psychol.* 512 (1996).

⁶¹ See Marilyn B. Brewer & Rupert J. Brown, *Intergroup Relations*, in 2 *The Handbook of Social Psychology* 554 (Daniel T. Gilbert et al. eds. 1998); Anthony G. Greenwald et al., *Implicit Partisanship: Taking Sides for No Reason*, 83 *J. Personality*

treatment of ingroup bias primarily as a cognitive phenomenon, rather than a motivated one.⁶² The familiarity explanation, however, fails to take account of research showing that people align themselves with people in their groups once they become aware of their commonality, even if they do not actually know members of the group or share meaningful experiences with them. Ingroup favoritism occurs in experimental settings when group designation is arbitrary, random, or meaningless.⁶³ Indeed, simply knowing that there are people outside the group can be enough to spur intergroup bias and rivalry—again, even when there is no contact with or knowledge of the others.⁶⁴ Based on these findings, most researchers have concluded that familiarity itself does not explain ingroup bias in these “minimal group” settings.⁶⁵

& Soc. Psychol. 367, 369 (2002). “Homophily,” or the attraction of people toward others like themselves, has been explained under relational demography theory as an “uncertainty reduction mechanism that facilitates individuals’ decisions regarding group membership and their formations of personal identities.” See Christine M. Riordan et al., *Relational Demography Within Groups: Through the Lens of Discrimination, in Discrimination at Work: The Psychological and Organizational Bases* 37, 39 (Robert L. Dipboye & Adrienne Colella eds., 2005).

⁶² Krieger, *supra* note 7, at 1191–93.

⁶³ Leslie Ashburn-Nardo et al., *Implicit Associations as the Seeds of Intergroup Bias: How Easily Do They Take Root?*, 81 *J. Personality & Soc. Psychol.* 789, 793 (2001) (finding ingroup bias even when groups are formed by whether their assigned name has an “X” or a “Q” in it); Henri Tajfel et al., *Social Categorization and Intergroup Behavior*, 1 *Eur. J. Soc. Psychol.* 149, 154–55 (1971) (finding that people exhibit ingroup favoritism even when groups formed according to whether people overestimate or underestimate the sizes of dots); see also Nilanjana Dasgupta et al., *Automatic Preference for White Americans: Eliminating the Familiarity Explanation*, 36 *J. Exper. Soc. Psychol.* 316 (2000) (reviewing study about the preference for familiar names). Anthony Greenwald puts ingroup favoritism in the context of a larger phenomenon whereby people are invited to “take sides,” even when the side they take is as arbitrary as home town, or which team is currently winning, or losing, a game. See Greenwald et al., *supra* note 61, at 367 (describing a “natural proclivity for partisanship”).

⁶⁴ This was demonstrated dramatically in the Robbers Cave experiments by Muzafer Sherif et al., *Intergroup Conflict and Cooperation: The Robbers Cave Experiment* (1961).

For an account in the social norms literature of how groups both use intra-group status rewards as a non-material means of gaining material sacrifice from members and how this dynamic leads to conflict between groups (and helps to explain racial discrimination), see Richard H. McAdams, *Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination*, 108 *Harv. L. Rev.* 1003, 1007–08 (1995).

⁶⁵ See, e.g., Ashburn-Nardo et al., *supra* note 63, at 795.

The fuller explanation for ingroup bias is that it serves people's psychological needs—in particular, the need to think well of themselves. According to social identity theory, people attribute positive characteristics to groups of which they are members in order to maintain a positive self-identity and thus heighten self-esteem.⁶⁶ The link between ingroup bias and self-esteem is supported by research demonstrating that threatening circumstances increase ingroup favoritism and outgroup disparagement, as people seek to shore up their sense of well-being and worth. Threats of all varieties have been associated with intolerance, especially threats to self-esteem.⁶⁷ In one set of studies, student test subjects showed an elevation in race stereotyping after receiving negative feedback from a black manager.⁶⁸ The effect was observed only when the student was on the receiving end of the feedback; merely witnessing negative feedback being given to another did not increase stereotyping.⁶⁹ Other research supports the conclusion that this type of response is not simply a generalized reaction to negative feedback, but a specific phenomenon related to stereotyping. In

⁶⁶ See Henri Tajfel & John C. Turner, *The Social Identity Theory of Intergroup Behavior Relation*, in *Psychology of Intergroup Relations* 7, 16 (Stephen Worchel & William G. Austin eds., 1986). For a description of similar theories to explain ingroup bias in relation to self-esteem and group identity, see Anthony G. Greenwald et al., *A Unified Theory of Implicit Attitudes, Stereotypes, Self-Esteem, and Self-Concept*, 109 *Psychol. Rev.* 3, 9–10 (2002). See also John C. Turner & Rina S. Onorato, *Social Identity, Personality, and the Self-Concept: A Self-Categorization Perspective*, in *The Psychology of the Social Self* 11 (Tom Tyler et al. eds., 1999).

⁶⁷ See Crocker et al., *supra* note 5, at 508 (threats to people's self-esteem increases ingroup favoritism); Stanley Feldman & Karen Stenner, *Perceived Threat and Authoritarianism*, 18 *Pol. Psychol.* 741, 762 (1997) (threats are associated with intolerance, especially among those with authoritarian attitudes); David A. Wilder & Peter N. Shapiro, *The Role of Competition-Induced Anxiety in Limiting the Beneficial Impact of Positive Behavior by an Out-Group Member*, 57 *J. Personality & Soc. Psychol.* 60, 67 (1989) (anxiety increases stereotyping of outgroup members, and reduced anxiety increases positive views of outgroup members).

⁶⁸ Lisa Sinclair & Ziva Kunda, *Reactions to a Black Professional: Motivated Inhibition and Activation of Conflicting Stereotypes*, 77 *J. Personality & Soc. Psychol.* 885, 887–94 (1999). Conversely, positive feedback from the same black manager reduced black stereotypes. *Id.* In follow-up studies, positive feedback from a black doctor appeared to trigger doctor stereotypes, whereas negative feedback triggered negative black stereotypes. *Id.* at 987–98; see also Spencer et al., *Automatic Activation of Stereotypes: The Role of Self-Image Threat*, 24 *Personality & Soc. Psychol. Bull.* 1139 (1998) (finding that negative feedback by Asian Americans increased activation of Asian-American stereotypes).

⁶⁹ Sinclair & Kunda, *Reactions to a Black Professional*, *supra* note 68, at 901–02.

one study, for example, students rated women instructors from whom they had gotten low grades as less competent to a substantially greater extent than male instructors from whom they had gotten low grades.⁷⁰ Consistent with the black manager feedback studies, students receiving high grades did not view their women instructors as any less competent than their male instructors.⁷¹ Negative feedback may trigger stereotypes even when it is not connected to the source of the feedback. In one experiment, (bogus) negative computer feedback on an intelligence test appeared to trigger stereotypes about gays and Jews, unrelated to the feedback or its source, that were not triggered when the feedback was positive.⁷²

The negative stereotyping produced in these types of experiments is typically inconsistent with what people report about their more egalitarian attitudes and beliefs about members of other groups. In an effort to explain the discrepancy, researchers have identified phenomena they label “aversive” racism⁷³ and “aversive” or “ambivalent” sexism.⁷⁴ Aversive racists think of themselves as egalitarian. They report their own views as egalitarian, sympathize with blacks, and are cautious about expressing negative evaluations. In situations in which they can do so without conscious awareness or exposing their true attitudes to others, however, they engage in stereotyping, reflect negative attitudes about racial minorities, and exhibit anxiety about interracial interaction.⁷⁵ Am-

⁷⁰ Lisa Sinclair & Ziva Kunda, *Motivated Stereotyping of Women: She's Fine if She Praised Me But Incompetent if She Criticized Me*, 26 *Personality & Soc. Psychol. Bull.* 1329, 1332–33 (2000).

⁷¹ *Id.* The findings of the study were replicated in a follow-up study of feedback by manager in a simulated work setting. *Id.* at 1334–37. Again, the effect of negative feedback on manager evaluation was not observed when the study subject merely observed the manager giving negative feedback to others. *Id.* at 1338–40.

⁷² Steven Fein & Steven J. Spencer, *Prejudice as Self-Image Maintenance: Affirming the Self Through Derogating Others*, 73 *J. Personality & Soc. Psychol.* 31, 32–37 (1997).

⁷³ The term “aversive racism” was first coined by Joel Kovel, see Kovel, *supra* note 23, at 54–55 (1970). It was revised and popularized principally through the work of John F. Dovidio and Samuel L. Gaertner. See, e.g., John F. Dovidio & Samuel L. Gaertner, *Aversive Racism and Selection Decisions: 1989 and 1999*, 11 *Psychol. Sci.* 315 (2000); Gaertner et al., *Aversive Racism*, *supra* note 23.

⁷⁴ Crisp & Turner, *supra* note 23, at 166–68; Glick & Fiske, *The Ambivalent Sexism Inventory*, *supra* note 23, at 491.

⁷⁵ See Gaertner et al., *Aversive Racism*, *supra* note 23, at 377–85.

2009]

Making Good on Good Intentions

1917

bivalent sexism works somewhat differently. Ambivalent sexists hold a mixture of negative and positive views toward women. For example, the belief that women are less competent than men at work, weaker, and less rational, but also that they are also warmer, more sensitive to the needs of others, and better at forming relationships. The positive or benevolent views appear to satisfy people's need to think of themselves as egalitarian, while at the same time helping them to rationalize or obscure their more negative or patronizing views.⁷⁶

The first experiments documenting the existence of aversive racism suggested that whites were less likely to give help to blacks than to other whites, unless circumstances made clear to the white test subjects that they were the only ones available to help. The apparent explanation was that whites harbored negative attitudes toward blacks that they were able to suppress when failing to do so would reveal those attitudes to themselves.⁷⁷

One set of studies from the early 1970s tested the responses of people to a phone call from a motorist who claimed to have broken down on the highway and, having used his last change trying to phone a service station and getting a "wrong number" instead, requested that the person on the other end of the line call the service station for him. Whites were less likely to help black callers (identifiable, researchers noted, by their dialects) than other whites. In other findings, people presumed by their party affiliation to be liberal treated black and white callers more equally than people presumed to be conservatives (in large part because conservatives helped whites more than liberals did). Liberals also expressed more egalitarian views. Liberals, however, also tended to try harder than conservatives to avoid learning information that might make clear that their help was needed—that is, they hung up the telephone earlier and more often.⁷⁸ This finding suggested to re-

⁷⁶ Glick & Fiske, *supra* note 23, at 494, 508. Researchers have noted that negative attitudes toward women have always been more ambivalent than negative racial attitudes, because of the "subjectively positive feelings toward women that often go hand in hand with sexist antipathy." *Id.* at 491.

⁷⁷ These experiments are summarized in Gaertner et al., *supra* note 23, at 381–83.

⁷⁸ See Samuel L. Gaertner, *Helping Behavior and Racial Discrimination Among Liberals and Conservatives*, 25 *J. Personality & Soc. Psychol.* 335 (1973). In this "wrong-number" study, each caller tried to explain that his or her car was broken down and that he or she was trying to reach a service station from a parkway tele-

searchers that people with views that are more consciously egalitarian nonetheless manifest negative attitudes or anxieties in subtle ways, and find strategies to obscure these attitudes from themselves and others.⁷⁹

Subsequent studies, while documenting a decline in overall self-reported expressions of prejudice, have confirmed the ambivalence phenomenon. One 1999 study involves the evaluations of employment applications. In the study, test subjects evaluated blacks with strong applications somewhat more highly than similarly-qualified whites (recommending the black candidate 87 percent of the time, compared to 79 percent for the white candidate), and they evaluated blacks with weak applications somewhat less negatively (recommending the black candidate 15 percent of the time, as compared to 8 percent for the white candidate). These same people, however, evaluated blacks with mixed qualifications—that is, strong on one criterion and weak on another—more negatively than similarly qualified whites (recommending the black candidate 40 percent of the time, as compared to 77 percent for the white candidate).⁸⁰

Similar results were obtained in studies in which test subjects were asked to review college applications. As in the employment application studies, test subjects did not discriminate against black applicants when the credentials were consistently strong or weak. They rated applications from blacks more negatively than comparable applications from white candidates, however, when the credentials were mixed. In these closer, more ambiguous circumstances, test subjects would adjust the weights given to various criteria in order to justify their judgments in particular case. Thus, when reviewers justified a negative decision, they gave greater weight to grades when the grades were weaker, and greater weight

phone, but now, as a result of apparently dialing a wrong number, was out of change, and needed the person on the other end of the line to telephone the garage. *Id.*

⁷⁹ This view is consistent with more recent research in the gender context concluding that more negative treatment of women is apparent in low-conflict, low-visibility situations, where it is harder to characterize as discriminatory, and easier to rationalize. See Bongsoon Cho & Debra L. Connelley, *The Effect of Conflict and Power Differentials on Social Identity and Intergroup Discrimination* 15–17 (unpublished manuscript, available at <http://ssrn.com/abstract=320286>).

⁸⁰ See, e.g., John F. Dovidio & Samuel L. Gaertner, *Aversive Racism*, *supra* note 23, at 317.

2009]

Making Good on Good Intentions

1919

to college board scores or the number of AP classes when the disfavored candidate was weaker according to these criteria.⁸¹ Researchers have identified a similar phenomenon in the gender context. In one experimental study, men were favored over women by study participants evaluating resumes for a construction job, with education or experience rated as the more important criterion depending upon the gender and credentials of the candidate.⁸²

Research attempting to determine when women and minorities benefit from a leniency bias and when they are judged more critically suggests that leniency bias is more common in non-zero-sum circumstances—that is, when it matters least. People give more praise to a woman than a man for a base hit in softball,⁸³ and stronger feedback to a black employee for a memo than they would give to a memo of the same quality written by a white employee.⁸⁴ Researchers believe that this favoritism, when it occurs, is motivated by a desire not to appear biased, as well as a generalized sense of awkwardness or ambivalence in interacting with minorities and women.⁸⁵ Overcorrection or benevolence represents a form of “giving credit where credit is due,” or sympathy, affirming people’s

⁸¹ See Gordon Hodson et al., Processes in Racial Discrimination: Differential Weighting of Conflicting Information, 28 *Personality & Soc. Psychol. Bull.* 460, 469–70 (2002).

⁸² See Michael I. Norton et al., Casuistry and Social Category Bias, 87 *J. Personality & Soc. Psychol.* 817, 820–21 (2004). It appears that the qualifications of favored candidates were not only weighted in order to justify the end result, but information was also viewed initially and remembered so as to support the favoritism, even before evaluators knew they would be asked to make a selection. *Id.* at 824–25.

⁸³ Monica Biernat, Toward a Broader View of Social Stereotyping, 58 *Am. Psychologist* 1019, 1019 (2003).

⁸⁴ *Id.* at 1021. According to Biernat, this is more likely when subjective standards are used, and when non-zero-sum judgments are being made. *Id.* at 1024–25; see also Lee Jussim et al., The Nature of Stereotypes: A Comparison and Integration of Three Theories, 52 *J. Personality & Soc. Psychol.* 536, 541–42 (1987) (finding that in an experimental setting, blacks who spoke in Standard English and dressed like members of a high socio-economic class were evaluated more favorably in mock-hiring context than similar whites).

⁸⁵ See Kent D. Harber, Feedback to Minorities: Evidence of a Positive Bias, 74 *J. Personality & Soc. Psychol.* 622, 626–27 (1998). This motivation may even elicit more positive feedback to blacks who act in an unfriendly way. See Kent D. Harber, The Positive Feedback Bias as a Response to Out-Group Unfriendliness, 34 *J. Applied Soc. Psychol.* 2272 (2004).

1920

Virginia Law Review

[Vol. 95:1893]

sense of themselves as unprejudiced.⁸⁶ When, however, employers choose among people for a single available position—that is, when it matters most—judgments tend to revert to stereotype. The woman who receives high praise for the base hit is not necessarily chosen next time for the team, and while women and minorities may get higher performance reviews and even make the “short list” for promotions, they are less likely to get them than their white male counterparts.⁸⁷

These complexities pose a difficult challenge for employment discrimination law. The studies predict that implicit discrimination will occur in evaluation contexts, but they cannot tell in individual instances when it will occur, or what form it will take. The research also suggests that people can identify strongly with egalitarian values, and yet unconsciously hold prejudiced attitudes that influence their evaluative decisions. These attitudes appear to be, at least in part, motivated by people’s needs to feel good about themselves. These findings raise significant questions about the optimal role for law in reducing discrimination. Before addressing these questions, Part II describes how Title VII currently addresses implicit bias, and how some scholars propose to reform the law to make it more effective.

II. TITLE VII AND IMPLICIT BIAS

A. Title VII Requires Causation, Not Deliberate Intent

Title VII prohibits workplace discrimination “because of” race, sex, and other protected characteristics.⁸⁸ Consistent with this language, both supporters and opponents of Title VII understood from the beginning that the Act was intended to address subtle, as

⁸⁶ See John F. Dovidio et al., *Why Can’t We Just Get Along? Interpersonal Biases and Interracial Distrust*, 8 *Cultural Diversity & Ethnic Minority Psychol.* 99, 92, 98–99 (2002). For the observation that overcompensation of candidates appears to represent the recognition of bias, but a naïve form thereof, see Christopher L. Aberson & Tara E. Ettlín, *The Aversive Racism Paradigm and Responses Favoring African-Americans: Meta-Analytic Evidence of Two Types of Favoritism*, 17 *Soc. Just. Res.* 25, 40–41 (Mar. 2004).

⁸⁷ See Biernat, *supra* note 83, at 1024–25; Nina Gupta et al., *Employee Gender, Gender Similarity, and Supervisor-Subordinate Cross-Evaluations*, 8 *Psychol. Women Q.* 174, 179 (1983).

⁸⁸ 42 U.S.C. § 2000e-2(a)(1) (2006).

2009]

Making Good on Good Intentions

1921

well as overt, forms of discrimination.⁸⁹ To this end, the U.S. Supreme Court has developed a proof structure enabling plaintiffs to establish the necessary causal link between an employer decision and race, sex, or other prohibited characteristic, even in the absence of explicit discriminatory intent. Under a line of cases beginning with *McDonnell Douglas Corp. v. Green*,⁹⁰ once a plaintiff proves a prima facie case,⁹¹ the defendant bears the burden of producing a non-discriminatory explanation for a negative employment decision. Defendant's production shifts the burden back to the plaintiff, who has the opportunity to show that the employer's reason was a pretext. If this burden is met, the factfinder is permitted to infer that the employer based its decision on an illegal reason, although it is not required to do so.⁹² While throughout this burden shifting courts often speak of discriminatory motive and intention, they also identify as the ultimate objective the determination of whether the employment action was *because of* a prohibited factor.⁹³

⁸⁹ See Chad Derum & Karen Engle, *The Rise of the Personal Animosity Presumption in Title VII and the Return to "No Cause" Employment*, 81 *Tex. L. Rev.* 1177, 1195–1209 (2003) (examining legislative history).

⁹⁰ 411 U.S. 792 (1973); see also *Tex. Dep't of Cmty. Affairs v. Burdine*, 450 U.S. 248, 248 (1981) (explaining further the burden of proof set forth in *McDonnell Douglas Corp. v. Green*).

⁹¹ A prima facie case is established when plaintiff proves that he or she is a member of a class protected by Title VII, and failed to obtain an employment opportunity for which she was qualified, or was subjected to an employment action to which others were not subject. *Burdine*, 450 U.S. at 248.

⁹² *Reeves v. Sanderson Plumbing Prods., Inc.*, 530 U.S. 133, 148 (2000).

⁹³ See, e.g., *id.* at 141 ("intentional" used to mean whether the "protected trait . . . actually motivated the employer's decision" (quoting *Hazen Paper Co. v. Biggins*, 507 U.S. 604, 610 (1998))); *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 506–11 (1993) (referring interchangeably to the plaintiff's need to prove intentional discrimination, and to prove that race was the "true reason" for the employment decision); see also *Price Waterhouse v. Hopkins*, 490 U.S. 228, 294 (1989) (Kennedy, J., dissenting) (the "ultimate question" is "whether discrimination caused the plaintiff's harm"). For an analysis of the difference between acting in fact on the basis of race or gender, and intending to do so, see Deborah Hellman, *When Is Discrimination Wrong?*, 138–68 (2008).

For a more comprehensive argument that Title VII already prohibits unconscious, as well as conscious, bias, see Banks & Ford, (How) Does Unconscious Bias Matter?, *supra* note 14, at 1072–89. See also Wax, *Discrimination as Accident*, *supra* note 12. For fuller treatment of the confusion over the concept of intent, see Wax, *The Discriminating Mind*, *supra* note 12, at 982–83.

Many employment discrimination scholars claim that courts have read an intent requirement into Title VII and urge the elimination of this requirement.⁹⁴ This contention has some support in the cases, but the point is often exaggerated. While courts frequently use language of discriminatory intent and motive in disparate treatment cases, typically they do so in order to distinguish discrimination based on a prohibited factor from employment action taken for non-discriminatory reasons, such as a violation of company rules or a lack of qualifications. In this context, the language of intention means that a causal link must be found between an employment action and the plaintiff's race, sex, or other protected characteristic—not that a deliberately or consciously discriminatory purpose is required.⁹⁵ Courts also use the language of intentional discrimination to distinguish disparate treatment cases from cases of disparate impact.⁹⁶ The 1991 Amendments to Title

⁹⁴ See, e.g., Lu-in Wang, *Discrimination by Default: How Racism Becomes Routine* 19–23 (2006); Martha Chamallas, *Deepening the Legal Understanding of Bias: On Devaluation and Biased Prototypes*, 74 S. Cal. L. Rev. 747, 752 (2001); Tristin K. Green, *A Structural Approach as Antidiscrimination Mandate: Locating Employer Wrong*, 60 Vand. L. Rev. 849, 895–900 (2007); Krieger, *supra* note 7, at 1164, 1168–73; McGinley, *supra* note 17, at 417–18; Robert L. Nelson et al., *Divergent Paths: Conflicting Conceptions of Employment Discrimination in Law and the Social Sciences*, 4 Ann. Rev. L. & Soc. Sci. 103, 107 (2008); David Benjamin Oppenheimer, *Negligent Discrimination*, 141 U. Pa. L. Rev. 899–900 (1993); Russell K. Robinson, *Perceptual Segregation*, 109 Colum. L. Rev. 1093, 1102 (2008); see also Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 Cal. L. Rev. 969, 980 (2006) (identifying that the “central focus” of existing antidiscrimination law is to prohibit “consciously biased decisionmaking”); Stacy E. Seicshnaydre, *Is the Road to Disparate Impact Paved With Good Intentions?: Stuck on State of Mind in Antidiscrimination Law*, 42 Wake Forest L. Rev. 1141, 1143–44 (2007) (criticizing courts for importing intent requirements into disparate impact cases).

⁹⁵ See *supra* note 93. Thus, for example, an employer's decision not to promote a woman because of the employer's assumption that she would not want to relocate because of family obligations might be based on the employer's concern for the employee, but it still constitutes sex-based discrimination. Cf. *Lust v. Sealy*, 383 F.3d 580, 583 (7th Cir. 2004) (affirming trial verdict in favor of plaintiff after the supervisor admitted that he didn't consider recommending the plaintiff for promotion because she had children). See also *Back v. Hastings*, 365 F.3d 107 (2d Cir. 1999) (reversing summary judgment against school district that denied tenure to a school psychologist on the assumption that, as a young mother, she would not be able to devote herself to the job).

⁹⁶ See, e.g., *Int'l Bhd. Teamsters v. United States*, 431 U.S. 324, 335 n.15 (1977) (disparate treatment under Title VII, to which “discriminatory motive is critical,” means “the employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin,” as contrasted with disparate impact,

2009]

Making Good on Good Intentions

1923

VII explicitly use the term “intentional discrimination” in this way,⁹⁷ rather than to signify an intent requirement that goes beyond the need to show causation.⁹⁸ Accordingly, in interpreting the Amendments, the Supreme Court has included circumstances such as the employer not understanding that certain conduct was unlawful within the rubric “intentional” and thus sufficient to support a claim for compensatory damages, while requiring that “malice,” “reckless indifference,” or “evil intent” be shown in order to support an award for punitive damages.

Some scholars point to the “honest belief” rule as evidence that courts require conscious intent to prove a discrimination case.⁹⁹ Here, again, the claim is overstated. Some courts have exonerated defendants who honestly believed that they had a nondiscriminatory reason for an employment action, such as that an employee was stealing from the company.¹⁰⁰ Generally, however, the relevance of evidence of honest belief is not that it disproves discrimination itself, but that it may undermine a showing of pretext.¹⁰¹ As

which involves employer practices that are facially neutral but fall more harshly on one group than another). Disparate impact discrimination was first identified by the Supreme Court in *Griggs v. Duke Power Co.*, 401 U.S. 424, 431–32 (1971). It was codified as part of Title VII in the Civil Rights Act of 1991. See 42 U.S.C. § 2000e-2(k) (2006).

⁹⁷ See 42 U.S.C. § 1981a(a)(1) (2006). This section provides for an award of compensatory and punitive damages against a “respondent who engaged in unlawful intentional discrimination” which is then explained in parentheses immediately thereafter: “(not an employment practice that is unlawful because of its disparate impact).”

⁹⁸ *Kolstad v. American Dental Association*, 527 U.S. 526, 535–37 (1999).

⁹⁹ See Linda Hamilton Krieger & Susan T. Fiske, Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment, 94 Cal. L. Rev. 997, 1029–38 (2006) (criticizing the “honest belief” rule). Of similar, claimed effect is the “same actor” inference, which permits a factfinder to conclude that the supervisor who hired the plaintiff did not later discriminate on the basis of a characteristic that would have been evident at the time of hiring. See Natasha T. Martin, Immunity for Hire: How the Same-Actor Doctrine Sustains Discrimination in the Contemporary Workplace, 40 Conn. L. Rev. 1117, 1124–38 (2008) (criticizing the “same actor” rule); Julie S. Northup, The “Same Actor Inference” in Employment Discrimination Cases: Cheap Justice?, 73 Wash. L. Rev. 193 (1998) (same).

¹⁰⁰ See, e.g., *Forrester v. Rauland-Borg Corp.*, 453 F.3d 416, 417–18 (7th Cir. 2006) (holding that whether employer discriminated is determined by his honest reason for taking the action).

¹⁰¹ Examples of the application of the honest belief rule that are criticized for their insistence upon a conscious intent requirement but which are consistent with this analysis include *Johnson v. AT&T Corp.*, 422 F.3d 756, 762 (8th Cir. 2005); *Millbrook v. IBP, Inc.*, 280 F.3d 1169, 1175 (7th Cir. 2002), and *Hawkins v. PepsiCo, Inc.*, 203

Richard Ford points out, if the rule sometimes produces wrong results in the jurisdictions in which it is applied, it is less because of judicial misunderstandings about a legal requirement of conscious intent than because of their overly simplistic notions of causation.¹⁰² When Title VII is read properly, a plaintiff should be able to show that an employer discriminated against her unknowingly, notwithstanding the employer's absence of deceit.¹⁰³

The fact that a showing of conscious intent is not required of plaintiffs under Title VII is apparent from the wide range of fact situations in which discrimination has been found under Title VII without a showing of hostility or intentionality. For example, disparaging remarks or stereotypes based on race or sex, procedural irregularities, or excessive reliance on subjective job criteria¹⁰⁴ all have sufficed to establish unlawful discrimination in disparate treatment cases, even in the absence of deliberate or conscious discriminatory intent. Findings of discrimination based on such evidence demonstrate that, while sometimes discriminatory animus is the best way to establish the link between an employment action and a characteristic protected by Title VII (or "intent"), it is not the only way, or even the most common.¹⁰⁵

The development of the mixed-motives doctrine also demonstrates that courts understand the discrimination prohibited by Ti-

F.3d 274, 279–80 (4th Cir. 2000), cert. denied, 531 U.S. 875 (2000). For application of the same actor rule, see *Proud v. Stone*, 945 F.2d 796, 798 (4th Cir. 1991), and cases cited in Martin, *supra*, note 99.

¹⁰² Cf. Banks & Ford, *supra* note 14, at 1087 (honest belief rule "does not make liability contingent on whether defendant consciously believed it discriminated [but] on whether the plaintiff can show that the defendant's proffered rationale was a pretext").

¹⁰³ See *Ricci v. DeStefano*, 129 S. Ct. 2658, 2674–75 (2009) (employer's good-faith belief that its race-based actions were necessary to avoid disparate impact liability is not enough to justify taking adverse action "because of . . . race").

¹⁰⁴ See, e.g., *Price Waterhouse v. Hopkins*, 490 U.S. 228, 235 (1989) (sex stereotypes); *Garrett v. Hewlett-Packard*, 305 F.3d 1210, 1217 (10th Cir. 2002) (procedural irregularities); *McCullough v. Real Foods, Inc.*, 140 F.3d 1123, 1129 (9th Cir. 1998) (hiring persons with lower job qualifications than the plaintiff); *Simms v. Oklahoma*, 165 F.3d 1321, 1328 (10th Cir. 1999) (subjective standards); *Thomas v. Eastman Kodak Co.*, 183 F.3d 38, 58 (1st Cir. 1999) (subjective employment practices); see also *EEOC v. Inland Marine Indust.*, 729 F.2d 1229, 1236 (9th Cir. 1984) (recognizing that racial discrimination "sometimes wears a benign mask").

¹⁰⁵ See Michael J. Zimmer, *A Chain of Inferences Proving Discrimination*, 79 U. Colo. L. Rev. 1243, 1243 (2008).

2009]

Making Good on Good Intentions

1925

tle VII to be a complex phenomenon, and not always the result of a conscious or single-minded effort to treat racial minorities or women worse than others. In *Price Waterhouse v. Hopkins*,¹⁰⁶ evidence that sex stereotypes about the plaintiff entered into the decision to deny her promotion to firm partnership was mixed with arguably sex-neutral evaluations of her harsh personality and difficulty in getting along with others. There was no evidence of a deliberate decision by the company partners to discriminate against her because of her sex. Nevertheless, the Court recognized that discrimination prohibited by Title VII can be the result of a mixture of motives, none of them necessarily intentional in the sense of deliberate or conscious. Members of the Court split over whether plaintiffs should have to prove that sex was a “but-for” cause of the action,¹⁰⁷ a “substantial factor,”¹⁰⁸ or a “motivating part,”¹⁰⁹ but all Justices appeared to agree that decisionmaking based on gender stereotypes violates Title VII, whether or not the wrongdoer is aware of the stereotypes or intends to engage in them.¹¹⁰ When Congress strengthened the mixed-motives standard in the Civil Rights Act of 1991, its choice of the “motivating factor” standard carried forward the *Price Waterhouse* meaning of the term “motive” as causation, not conscious or deliberate intent.¹¹¹ In later holding that mixed-motive cases could be shown by circumstantial as well as direct evidence,¹¹² the Court made still clearer that there are many ways to prove that discrimination has oc-

¹⁰⁶ 490 U.S. 228 (1989).

¹⁰⁷ *Id.* at 295 (Kennedy, J., dissenting).

¹⁰⁸ *Id.* at 276–77 (O’Connor, J., concurring).

¹⁰⁹ *Id.* at 250 (plurality opinion).

¹¹⁰ Thus, even the dissenting opinion written by Justice Kennedy and joined by Chief Justice Rehnquist and Justice Scalia, while incorporating the language of discriminatory intent, states that the “ultimate question . . . is whether discrimination caused the plaintiff’s harm.” *Id.* at 294 (Kennedy, J., dissenting); see also *id.* at 295.

¹¹¹ The Court in *Price Waterhouse* held that the employer was not liable if it could show that it would have reached the “same decision” on the basis of nondiscriminatory factors. *Id.* at 242. The Civil Rights Act of 1991 changed the same-decision rule, providing that liability is established under Title VII if any prohibited factor motivated the decision, see 42 U.S.C. § 2000e-2(m) (2006), but limiting damages in cases in which the employer shows that its decision was justified on the basis of a nondiscriminatory factor. See 42 U.S.C. § 2000e-5(g)(2)(B) (2006).

¹¹² *Desert Palace, Inc. v. Costa*, 539 U.S. 90, 100 (2003).

curred, and that there is no set formula requiring deliberate intent.¹¹³

These cases affirm that Title VII's prohibition of discrimination "because of" a plaintiff's race, sex, or other protected characteristic covers unconscious as well as conscious discrimination. In addition, courts' understanding of what discrimination means have encompassed increasingly more subtle forms of discrimination, proved by more indirect forms of proof. When the Supreme Court has renegeed on the promise of Title VII in this regard,¹¹⁴ it often has self-corrected,¹¹⁵ or Congress has intervened to repair the damage.¹¹⁶ As discussed later in the Article,¹¹⁷ further clarification that Title VII prohibits discrimination, not malicious intent, would correct some misunderstandings about how Title VII should be applied. It should be understood, however, that these reforms will not turn Title VII from a statute that fails to recognize unconscious bias to one that does, and thus are unlikely to have the hoped-for, major impact in reducing implicit discrimination.

B. Proposed Reforms

In addition to advocating movement away from the supposed intent requirement of Title VII, employment discrimination scholars have proposed various reforms that, in small and large ways, re-

¹¹³ Id. ("Circumstantial evidence is not only sufficient, but may also be more certain, satisfying and persuasive than direct evidence" (quoting *Rogers v. Missouri Pac. R.R. Co.*, 352 U.S. 500, 508 n. 17 (1957)) (citation omitted)).

¹¹⁴ See, e.g., *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 518–19, 523–24 (1993) (holding that proof of pretext requires more than discrediting defendant's "nondiscriminatory" reason); *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 659–60 (1989) (weakening business necessity standard in disparate impact cases, and shifting burden of persuasion on the issue to plaintiff); see also *Price Waterhouse*, 490 U.S. at 262–63, 276, 278 (O'Connor, J., concurring) (asserting that in mixed motives case, plaintiff must prove that sex was the "but for" cause, using direct evidence).

¹¹⁵ See, e.g., *Desert Palace, Inc.*, 539 U.S. at 92 (holding that direct evidence is not required in mixed motives case); *Reeves v. Sanderson Plumbing Prods., Inc.*, 530 U.S. 133, 147–49 (2000) (clarifying *Hicks*, 509 U.S. 502, to make clear that the factfinder is entitled to find for the plaintiff on the basis of the plaintiff's discrediting of the defendant's nondiscriminatory reason).

¹¹⁶ See, e.g., 42 U.S.C. § 2000e-2(k)(1)(A) (2006) (codifying *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), after the Supreme Court had cut back on the decision in *Wards Cove*, 490 U.S. 642); 42 U.S.C. § 2000e-2(m) (2006) (establishing that plaintiff can make out mixed motives case by showing that sex was a "motivating factor").

¹¹⁷ See *infra* Section V.A.

2009]

Making Good on Good Intentions

1927

spond to the difficulty of proving implicit discrimination. These proposals have in common a general sense that implicit discrimination is too hard to prove under current law and thus that this law should be deliberately recalibrated to bring about more positive results for plaintiffs. They also share an assessment that, given the difficulties of proof, the risk of false negatives outweighs the risk of false positives.¹¹⁸

Some of the proposals would alter the proof structure of Title VII and the burden of proof each party must carry. Ann McGinley, for example, argues in favor of a *mandatory* presumption of discrimination after the plaintiff makes a *prima facie* case of discrimination under *McDonnell Douglas/Burdine* and then demonstrates that the reason the defendant gives for its employment decision is inaccurate or untrue (that is, a “pretext”).¹¹⁹ Current law permits an inference of discrimination based on pretext, on the assumption that pretext may mean that the employer was lying in order to cover up a discriminatory action, but it does not require that conclusion.¹²⁰ The approach of current law recognizes, for example, that an employer who uses the recession as an excuse to let go some of its less productive employees might have tried to obscure the real cause of an employment action, but has not necessarily engaged in race or gender discrimination. The employee may have wished to avoid having to prove the performance problems, or the company may have desired to avoid unrelated, damaging public disclosure of the real motive, such as embezzlement, falling profits, or a sexual assault by the plaintiff of an innocent employee. The proposed presumption would not allow these alternatives to be considered.¹²¹

Another proposal would create a new liability category. Russell Robinson advocates the imposition of liability under Title VII when the discrimination is clear to those with backgrounds and experiences similar to those of the plaintiff, even though it is not clear to others. Robinson’s proposal is grounded in a theory of percep-

¹¹⁸ See McGinley, *supra* note 17, at 482 (“[T]he opportunities for underestimating the incidence of discriminatory behavior are much greater than the reverse.”).

¹¹⁹ *Id.* at 481–82.

¹²⁰ *Reeves*, 530 U.S. at 147–49; *Hicks*, 509 U.S. at 518–19, 523–24.

¹²¹ For other examples of proposals designed to alter the proof structure to make it easier to prove implicit discrimination, see Ivan E. Bodensteiner, *The Implications of Psychological Research Related to Unconscious Discrimination and Implicit Bias in Proving Intentional Discrimination*, 73 *Mo. L. Rev.* 83, 108, 120–27 (2008).

tual differences between insiders and outsiders, and the unfairness of limiting Title VII liability to circumstances recognizable by those whose insider perspective blinds them to many of the kind of discrimination outsiders experience.¹²² He would impose the new form of intermediate liability when “a reasonable outsider would find the claim compelling, although an insider judge might not.”¹²³ In such cases, plaintiffs could recover some, but not all, of the relief that they would obtain in a case in which they were able to prove discrimination that both insiders and outsiders would recognize.¹²⁴

Other commentators have proposed measures that would expand employer responsibility to correct whatever workplace procedures and structures might fail to block discriminatory decision-making. For example, David Oppenheimer suggests that Title VII should include a tort-like duty on the part of the employer to take all reasonable, affirmative precautions against discrimination.¹²⁵ While under current Title VII law employer practices and procedures that have a disparate impact on women or minorities are subject to challenge,¹²⁶ under Oppenheimer’s proposal, *every* rejection of a woman or minority job applicant would trigger a duty to stop and examine the motives of the decisionmaking instantly. In addition, Oppenheimer would impose liability on employers for failing to scrutinize job screening procedures, employee evaluations, and employee disciplinary actions to ensure they do not allow discrimination to improperly influence decisionmaking.¹²⁷ This approach, in essence, would mandate a uniform, albeit shifting, set of diversity “best practices” for every employer subject to Title VII.

Tristin Green, too, would hold employers responsible for correcting a broad range of “structural features” of the workplace that

¹²² Robinson sees this focus on perceptual differences between insiders and outsiders as an important and overlooked part of the implicit bias problem. See Russell K. Robinson, Perceptual Segregation, 108 Colum. L. Rev. 1093, 1103 (2008) (noting that implicit bias scholarship is about the psychology of privileged groups or insiders, which speaks to only part of the story of implicit bias).

¹²³ Id. at 1167.

¹²⁴ Id. at 1169 (plaintiff might receive “some form of relief, such as attorneys’ fees or a portion of the damages that a prevailing plaintiff would normally recover”).

¹²⁵ See Oppenheimer, *supra* note 94, at 899 (1993).

¹²⁶ See *Griggs v. Duke Power Co.*, 401 U.S. 424, 429–31 (1971).

¹²⁷ Oppenheimer, *supra* note 94, at 970.

might permit unconscious bias to occur.¹²⁸ Under Green's proposal, the failure to achieve demographic balance in work team assignments that might reduce bias could constitute a Title VII violation, as could the failure to reduce the salience of ingroup-outgroup distinctions, communication systems that do not achieve optimal information flow that could increase inclusiveness, and chains of command that reproduce race and gender hierarchies.¹²⁹ Employers would be expected to know how implicit bias operates, and to "refrain from creating work environments that facilitate the operation of those biases in workplace decisionmaking."¹³⁰ This responsibility would extend to anticipating, and taking steps to avoid, the responses that victims sometimes make to discrimination that further disadvantages them.¹³¹

These kinds of proposals represent understandable frustrations with a legal system that has not yet been able to get an effective grip on implicit discrimination. They assume that expanding liability standards and procedural rules making discrimination easier to prove will incentivize employers to be more careful about the deci-

¹²⁸ See, e.g., Green, *A Structural Approach as Antidiscrimination Mandate: Locating Employer Wrong*, supra note 18; see also Tristin K. Green & Alexandra Kalev, *Discrimination-Reducing Measures at the Relational Level*, 59 *Hastings L.J.* 1435, 1457 (2008) (suggesting that the law should require employers to "address relational sources of discrimination" and eliminate "extreme rigidity and segregation in job categories").

¹²⁹ Green, *A Structural Approach as Antidiscrimination Mandate: Locating Employer Wrong*, supra note 18, at 856–57.

¹³⁰ *Id.* at 899. Susan Sturm has also argued in favor of structural workplace reform. By and large, however, Sturm focuses not on enhanced liability rules for individual plaintiffs but on other kinds of workplace accountability and institutional change. See supra note 13.

¹³¹ Green, *A Structural Approach as Antidiscrimination Mandate: Locating Employer Wrong*, supra note 18, at 900–03. Green uses the term "accommodation mandate" to refer to employer responsibilities that stem from the obligation to anticipate the response of women and members of minority groups to existing workplace structures and to alter those structures to accommodate—and possibly interrupt—the cycle of bias. *Id.* at 902. The examples she gives of responsive behavior which the employer should anticipate are members of minority groups undertaking "extra identity work" to persuade others that they do not have the stereotyped characteristics attributed to them, and the lack of female interest in certain jobs which is the result of employer policies that have depressed that interest. *Id.* at 900–02; see also Tristin K. Green, *Discomfort at Work: Workplace Assimilation Demands and the Contact Hypothesis*, 86 *N.C. L. Rev.* 379, 379–80 (2008) (arguing that employers should be required to permit people of color and women to "signal" group identification by accommodating their dress and appearance needs).

sions they make. The fact that law has not been entirely effective in ridding the workplace of discrimination, however, does not mean that these wider and deeper legal nets would do any better. Specifically, proposals to short-circuit basic issues of proof and causation and second-guess a wide range of employer decisions ignore the potential negative effects of law on the very discrimination they seek to reduce. While these reforms may close off some opportunities for discrimination, they also risk undermining the conditions necessary to motivate people to want to avoid implicit discrimination. They may, in fact, feed rather than reduce implicit discrimination.

To demonstrate how this might be the case, Part III examines the principles of motivation and how they relate to whether people respond to others in nondiscriminatory ways. It examines the benefits of people internalizing the law's goals, and reviews the social science literature relevant to what supports, and what defeats, that internalization process. In light of those principles, Part IV then reviews social science findings relating specifically to the reduction of bias. Part V concludes by applying these findings to the workplace and to employment discrimination law.

III. PEOPLE'S MOTIVATIONS TO COMPLY WITH NONDISCRIMINATION NORMS

People are as differently motivated to comply with nondiscrimination norms as they are to comply with other laws and expectations. The difference is not just a question of degree or strength, but also of source or quality. There are two basic sources of motivation: external and internal. External motivation comes from the desire for approval or rewards or to avoid negative sanctions. Internal motivation comes from within the individual's personal values and identity structure. People respond to both, often at the same time, but as between the two, internal motivation has certain advantages over external motivation. Part A, below, addresses these advantages. Part B explores the role of law in building, or undermining, internal norms, and ties these principles directly to nondiscrimination norms.

2009]

Making Good on Good Intentions

1931

A. Distinguishing External and Internal Motivation

People are motivated by both external and internal pressure. Commands of an outside authority, such as a law or court order, motivate externally. People obey those commands to avoid negative consequences, such as legal penalties or fines. Social norms function in much the same way. People try to conform to those norms in order to gain approval (and avoid disapproval) by family, friends, co-workers, and other social networks.¹³² Approval matters to people because it affects self-image and self-esteem.¹³³ People who might otherwise feel no internal qualms about racist jokes will be less likely to tell them if they know that their friends disapprove, just as they will be more likely to recycle or refrain from littering if others whose views matter will think better of them as a result.¹³⁴ Rewards serve as external motivators, as do threats. A student who is paid for every “A” she receives, or penalized when she does poorly, may study harder in order to obtain top grades.

Internal, or intrinsic, motivation comes from people’s own sense of who they are and what they value. People who play the piano out of the pleasure they get from producing the sound rather than pleasing their teacher or their parents are internally motivated. People who work in a soup kitchen in order to be “true” to themselves¹³⁵ are living out their self-identity. So are people who try to make unprejudiced judgments about others because it is the right thing to do, rather than something the law compels.

The norms that people internalize typically derive from external rules and social norms. The internalization process, however, is

¹³² For a review of social norms literature, see Richard H. McAdams, *The Origin, Development, and Regulation of Norms*, 96 Mich. L. Rev. 338, 339–50 (1997).

¹³³ See *supra* text accompanying notes 66–72. One of the ways law motivates changed behavior, McAdams writes, is to “signal[] on what grounds the majority will henceforth give and withhold esteem.” Richard H. McAdams, *Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination*, 108 Harv. L. Rev. 1003, 1081 (1995).

¹³⁴ See generally Robert E. Scott, *The Limits of Behavioral Theories of Law and Social Norms*, 86 Va. L. Rev. 1603 (2000) (examining theories explaining why people respond positively to legal rules and social norms, and cautioning against non-falsifiable assumptions that rules and norms actually change preferences, as opposed to simply incentivize behavior).

¹³⁵ See Wendy Wood, *Motives and Modes of Processing in the Social Influence of Groups*, in *Dual-Process Theories*, *supra* note 19, at 547, 553 (identifying interest of being true to oneself separate from interest in gaining approval of social group).

neither inevitable nor automatic. Certain conditions encourage it, and others make it less likely. According to social psychologist Edward Deci, whether people internalize external norms is related to the extent to which they feel autonomous, competent, and related to others. The needs for autonomy, competence, and relatedness are “innate psychological nutriments that are essential for ongoing psychological growth, integrity, and well-being.”¹³⁶ These are distinct, though related needs. Autonomy refers to the desire to “self-organize experience and behavior and to have activity be concordant with one’s integrated sense of self.”¹³⁷ Competence is the sense of having “an effect on [one’s] environment” and a propensity “to attain valued outcomes within it.”¹³⁸ Relatedness refers to the desire “to love and care, and to be loved and cared for.”¹³⁹

The presence of these qualities affects a person’s self-esteem. As it relates to the capacity to internalize norms, self-esteem is measured, importantly, not only in terms of *how much* of it a person has, but also on the type of motivation in which it is grounded. Self-esteem based in the external need for approval is insecure. People with insecure self-esteem will tend to engage in behavior designed to satisfy their self-esteem needs, rather than other, more desirable social goals.¹⁴⁰ The more insecure their self-esteem, the more approval from others is required to maintain it. As a result, insecure self-esteem does not support the internalization of legal and social norms, such as nondiscrimination; it works against them.¹⁴¹ In contrast, self-esteem that is grounded in an internal sense of autonomy, competence, and connectedness does not require continuous external reinforcement; rather, it draws on internal norms that are *self-reinforcing*. In short, insecure self-esteem needs external verification; secure self-esteem perpetuates itself.

¹³⁶ Deci & Ryan, The “What” and “Why” of Goal Pursuits, *supra* note 20, at 227, 229 (2000).

¹³⁷ *Id.* at 231.

¹³⁸ *Id.* at 230–31.

¹³⁹ *Id.* at 231.

¹⁴⁰ *Id.* at 230–32.

¹⁴¹ See Patricia G. Devine et al., The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice, 82 *J. Personality & Soc. Psychol.* 835, 845 (2002) (noting that, in experimental studies, individuals with high levels of internal motivation and low levels of external motivation were more autonomous, and more effective in regulating race bias, even on tests involving responses that were difficult to control).

2009]

Making Good on Good Intentions

1933

The benefits of internalized norms and self-esteem are substantial. Research has shown that when they are internally motivated, people do better in protecting the environment, overcoming alcohol or drug addiction,¹⁴² losing weight, taking medications, and engaging in school or personal relationships.¹⁴³ They perform better especially with respect to performance that is difficult, or that requires effort and persistence.¹⁴⁴ They are also better able to recognize inconsistencies between their professed values and their actions, and to dampen their own stereotyped responses to others.¹⁴⁵

In the workplace, people who lack a sense of autonomy, competence, and relatedness and thus whose self-esteem is insecure will tend to go through the motions of a job without caring about the quality of the work product.¹⁴⁶ They will often engage in overly aggressive behaviors to gain approval that have the opposite effect.¹⁴⁷ Sometimes they create difficult social situations because of their insecurity or sense of disconnectedness and then respond to those situations by trying to save face, blaming others, or otherwise acting defensively.¹⁴⁸ These are the same characteristics associated with the lack of internally motivated goals, and with prejudice.¹⁴⁹ Even successful, powerful people can have self-esteem that is based on the need for more external affirmation, success, and authority. In fact, powerful people often depend the most on others

¹⁴² Deci & Ryan, The “What” and “Why” of Goal Pursuits, *supra* note 20, at 241 (citing studies).

¹⁴³ Richard M. Ryan & Edward L. Deci, Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being, 55 *Am. Psychologist* 68, 73 (2000) (citing studies).

¹⁴⁴ Deci & Ryan, The “What” and “Why” of Goal Pursuits, *supra* note 20, at 241, 243, 244.

¹⁴⁵ See John F. Dovidio et al., Reducing Contemporary Prejudice: Combating Explicit and Implicit Bias at the Individual and Intergroup Level, *in* *Reducing Prejudice and Discrimination* 137, 144–47 (Stuart Oskamp ed., 2000) (discussing evidence of ability of people who are internally motivated—or have “good intentions”—to regulate their own implicit biases).

¹⁴⁶ Ryan & Deci, Self-Determination Theory, *supra* note 143, at 72.

¹⁴⁷ Jennifer K. Bosson et al., Self-Enhancement Tendencies Among People With High Explicit Self-Esteem: The Moderating Role of Implicit Self-Esteem, 2 *Self & Identity* 169, 181 (2003).

¹⁴⁸ Deci & Ryan, The “What” and “Why” of Goal Pursuits, *supra* note 20, at 251.

¹⁴⁹ See E. Ashby Plant & Patricia G. Devine, Internal and External Motivation to Respond Without Prejudice, 75 *J. Personality & Soc. Psychol.* 811, 817–18, 826–27 (1998).

1934

Virginia Law Review

[Vol. 95:1893]

for their sense of self-worth. Accordingly, those who need to dominate situations as a basis of self-worth tend to form more biased judgments of others than do those with the same amount of power who do not depend upon control over others to sustain their sense of self.¹⁵⁰

B. The Role of Law in Building, or Dissipating, Internal Motivation to Reduce Implicit Bias

1. The Positive Role of Law on Norm Internalization

While the law operates in the first instance as an external motivator that defines wrongful behaviors and penalizes people who engage in them, it also affects people's internal norm commitments. First, the law influences people's attitudes and beliefs by giving them information about the attitudes and beliefs of others.¹⁵¹ Richard McAdams observes that, in the context of discrimination, law can help people to see that the case for discrimination is no longer compelling. "Rationalizations can be fragile things," McAdams writes, and the law can "symboliz[e] a consensus that the rationalizations for [discrimination] are, in fact, mere rationalizations."¹⁵² The law can stretch people's thinking, challenging them to think critically about and perhaps revise their thoughts, including judgments, decisions, and behaviors that may have been based on group stereotypes.¹⁵³

To the extent the law is successful in its role as external rule-definer, it also influences internal norms through the routines and processes it encourages, which structure the world in which people

¹⁵⁰ See Don Operario & Susan T. Fiske, Effects of Trait Dominance on Powerholders' Judgments of Subordinates, 19 *Soc. Cognition* 161, 176-77 (2001); see also Stephanie A. Goodwin et al., Power Can Bias Impression Processes: Stereotyping Subordinates by Default and by Design, 3 *Group Processes Intergroup Relations* 227, 251 (2000) (stereotyping of subordinates by people with power explained by the motivational pressures to maintain control, rather than by inattention, cognitive load, or disinterest).

¹⁵¹ Richard McAdams refers to people's mistaken beliefs about other people's attitudes as "pluralistic ignorance." McAdams, An Attitudinal Theory of Expressive Law, 79 *Or. L. Rev.* 339, 356-57 (2000).

¹⁵² Richard H. McAdams, Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination, 108 *Harv. L. Rev.* 1003, 1081 (1995).

¹⁵³ See Gregory Mitchell, Second Thoughts, 40 *McGeorge L. Rev.* 687, 689 (2009).

2009]

Making Good on Good Intentions

1935

interact with one another and thereby form beliefs and attitudes about each other.¹⁵⁴ When employers institute hiring and promotion procedures that put a premium on nondiscriminatory decisionmaking, these procedures accustom people to accounting for their hiring and promotion decisions in ways that reinforce nondiscriminatory norms. While conforming to the norms of a group may be prompted initially by the external pressure to conform, over time efforts to identify with the group help to transform internal attitudes and commitments as well.¹⁵⁵

Compliance with the law also reinforces a self-identity consistent with that behavior. Acting in a certain way contributes to people becoming people who act in a certain way. Even though I begin to recycle because the law requires me to do so or because I want to impress my friends, once I do so I become, in my own eyes and in the self I project to others, a person who recycles. This self-identification reinforces my recycling habits. Conversely, littering makes me a litterer, meaning someone who is more likely to litter in the future.¹⁵⁶

Right behavior, in turn, educates *others* about what it means to be a good person. In this sense, law creates what Robert George calls a better “moral ecology,” in which all people make their morally self-constituting choices.¹⁵⁷ In a workplace, when some people appear to follow nondiscrimination norms, this compliance feeds an environment that shapes other people’s sense of how they should act if they want to consider themselves good workplace citizens.

¹⁵⁴ For a careful examination of how habit moves from rote action to a conceptually integrated part of one’s self-identity, see Robin R. Vallacher & Daniel M. Wegner, What Do People Think They’re Doing? Action Identification and Human Behavior, 94 *Psychol. Rev.* 3, 4–13 (1987).

¹⁵⁵ See Christian S. Crandall et al., Social Norms and the Expression and Suppression of Prejudice: The Struggle for Internalization, 82 *J. Personality & Soc. Psychol.* 359, 375 (2002) (describing research suggesting that social norms can be internalized even by the individual who suppresses prejudice in order to conform to them).

¹⁵⁶ As put by Robert George, the law can prevent a person’s “self-corruption which follows from acting out a choice to indulge in immoral conduct.” Robert P. George, *Making Men Moral: Civil Liberties and Public Morality* 1 (1993).

¹⁵⁷ *Id.* For a discussion of how other people create the normative climate for people’s behavior, see Tom R. Tyler, *Why People Obey the Law* 24 (1990).

2. *The Negative Role of Law on Norm Internalization*

Despite these positive ways in which it strengthens public non-discrimination norms, the law does not always have a positive effect on the internalization of norms. There are also a number of potential negative effects. First, when people view the law as unfair or overly intrusive, they tend to resist it. As a result, even when they comply in order to avoid the negative consequences, they will not internalize the law's norms, or obey them when they can avoid doing so.¹⁵⁸ Research experiments support the conclusion that people tend to react to mandates they perceive as unfair by resisting the underlying goals of those mandates. In one study, for example, participants were leaned on by their "boss" at a food franchise to hire a black applicant for a job opening in order to increase the ethnic diversity of the company. As a result of the pressure, which consisted of a reminder that pay raises were to be decided in the next two weeks, test subjects made the decisions they were expected to make. Subsequently, however, they registered higher levels of anger, threat, and resentment than before, and greater opposition to affirmative action policies.¹⁵⁹

Similarly, in the corporate context, researchers have shown that workers tend to respond to superiors who have insulted their sense of identity—by means of surveillance systems, public criticism or other actions implying a lack of trust—by working less hard, bad mouthing others, or using company resources in an unauthorized

¹⁵⁸ See Tyler, *supra* note 157, at 4, 64–65, 165; see also Oliver Wendell Holmes, *The Common Law* 39 (The Belknap Press of Harvard University Press 2009) (1881) ("The first requirement of a sound body of law is, that it should correspond with the actual feelings and demands of the community, whether right or wrong."); Tyler, *supra*, at 25, 36–37, 64–65 (discussing how law is ineffective when it is not taken seriously by those who wish to be law-observant). Jack Knight describes the desired relation of law to existing beliefs and social norms as one of pragmatism. The idea is that law should be sufficiently rooted in society's existing beliefs so that it generates trust, and thus commands the buy-in necessary for it, in turn, to have further influence on society's beliefs. See Jack Knight, *Social Norms and the Rule of Law: Fostering Trust in a Socially Diverse Society*, in *Trust in Society* 354, 367–71 (Karen S. Cook ed., 2001).

¹⁵⁹ E. Ashby Plant & Patricia G. Devine, *Responses to Other-Imposed Pro-Black Pressure: Acceptance or Backlash?*, 37 *J. Experimental Soc. Psychol.* 486 (2001). This response was shown by study participants with a low internal motivation to avoid prejudice; those who started with a high level of motivation to avoid prejudice did not show the same elevation of resentment or the same backlash. *Id.* at 498–99.

2009]

Making Good on Good Intentions

1937

manner.¹⁶⁰ These responses have their own moral imperative; in one study, people who reported they would respond in these ways thought that they would be doing the right thing.¹⁶¹ These responses, in turn, create the need for more surveillance and tighter workplace controls.¹⁶²

Another way in which the law puts itself at odds with people's internalization of the law's underlying norms is the phenomenon of external pressure "crowding out" internal motivation. An external command signals to people that they are not expected to do the right thing on their own; they will do so only if compelled. This signal can undermine their sense of autonomy, competence, and relatedness, and elicit shame, resentment and resistance, "in ways that include abandoning self-regulation."¹⁶³

The "crowding out" phenomenon has been tested in a number of different contexts. In one well-known study, imposing a fine for late pickup of children from a daycare center increased, rather than decreased, late pickups. The theory is that whereas previously the parents felt an internal, moral obligation to pick up their children on time (although they did not always do so), the imposition of the fine commodified the activity, reducing the violation from a moral obligation to a mere matter of money. It informed parents that they are expected to pick up their children late from time to time,

¹⁶⁰ Robert J. Bies & Thomas M. Tripp, Beyond Distrust: "Getting Even" and the Need for Revenge, *in* Trust and Organizations 246, 251, 257 (Randall M. Kramer & Tom Tyler eds., 1996); Robert B. Cialdini, Social Influence and The Triple Tumor Structure of Organizational Dishonesty, *in* Codes of Conduct: Behavioral Research Into Business Ethics 44, 56-57 (David M. Messick & Ann E. Tenbrunsel eds., 1996); Edward L. Deci & Richard M. Ryan, The Support of Autonomy and the Control of Behavior, 53 *J. Personality & Soc. Psychol.* 1024, 1026 (1987); see also Ian Ayres & John Braithwaite, Responsive Regulation: Transcending the Deregulation Debate 53 (1992) (finding that if legal controls are too salient, corporate actors attribute their compliance to the law instead of their own ethical standards).

¹⁶¹ Bies & Tripp, *supra* note 160, at 258.

¹⁶² See Cialdini, *supra* note 160, at 57.

¹⁶³ Ayres & Braithwaite, *supra* note 160, at 25; see also Alice Erh-Soon Tay, Communist Visions, Communist Realities and the Role of Law, 17 *J. L. & Soc'y* 155, 159-60 (1990) ("If people be led by laws . . . and uniformity is sought to be given them by punishments, they will try to avoid punishments but have no sense of shame. If they be led by virtue . . . they will have a sense of shame and, moreover, will become good.") (quoting Confucius); see also Milton C. Regan, Jr., Risky Business, 94 *Geo. L.J.* 1957, 1958 (2006) (rules designed to make lawyers act ethically will not necessarily motivate them to act ethically).

and defined that option as a choice one might make (with certain liquidated costs), rather than as an unacceptable breach of an obligation.¹⁶⁴ They paid more, complied less, and felt less guilty about it.¹⁶⁵ Rewarding students for academic performance through such incentives as money, gold stars, and honor roles, likewise, may change the perceived nature of education from an inherently desirable undertaking that is worth doing well, to one that is not worth the effort unless the rewards are high enough.¹⁶⁶ In another context, one set of studies has shown that when payment is expected for a task, effort improves with the level of payment, but when no payment is expected, effort is as high as the effort made with the higher level of payment.¹⁶⁷ External incentives may also affect judgment. One commentator intuits that having an excessive num-

¹⁶⁴ This research is described in Uri Gneezy & Aldo Rustichini, *A Fine is a Price*, 29 *J. Legal Stud.* 1, 13–14 (2000). For citations to other research studies showing the “crowding-out” effect, see Yochai Benkler, *Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production*, 114 *Yale L.J.* 273, 321–26 (2004).

¹⁶⁵ Gneezy & Rustichini, *A Fine is a Price*, *supra* note 164, at 13–14. Conversely, being paid for something people might readily give for free may result in less giving. In the context of blood donations, see Richard M. Titmuss, *The Gift Relationship: From Human Blood to Social Policy* (1971); see also Carl Mellström & Magnus Johannesson, *Crowding Out in Blood Donation: Was Titmuss Right?*, 6 *J. Eur. Econ. Ass’n* 845 (2008) (discussing a Swedish experiment finding significant crowding-out effect of paying for blood among women but not among men).

¹⁶⁶ This point is hotly debated but supported by the strong weight of the evidence. Compare Judy Cameron & W. David Pierce, *Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis*, 64 *Rev. Educ. Res.* 363 (1994) (concluding that rewards only minimally decrease intrinsic motivation), and Judy Cameron & W. David Pierce, *The Debate About Rewards and Intrinsic Motivation: Protests and Accusations Do Not Alter the Results*, 66 *Rev. Educ. Res.* 39 (1996) (arguing that the questions, methods and techniques used in the authors’ earlier study were appropriate and that their conclusions were well-founded), with Edward L. Deci, *Effects of Externally Mediated Rewards on Intrinsic Motivation*, 18 *J. Personality & Soc. Psychol.* 105 (1971) (finding that when money was used as an external reward, intrinsic motivation tended to decrease, and when verbal reinforcement and positive feedback were used, intrinsic motivation tended to increase), and Edward L. Deci et al., *A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation*, 125 *Psychol. Bull.* 627 (1999) (concluding from a meta-analysis of 128 studies that rewards significantly undermine free-choice intrinsic motivation), and Edward L. Deci et al., *Extrinsic Reward and Intrinsic Motivation in Education: Reconsidered Once Again*, 71 *Rev. Educ. Res.* 1 (2001) (finding that more recent studies support the argument that external rewards decrease intrinsic motivation).

¹⁶⁷ See James Heyman & Dan Ariely, *Effort for Payment: A Tale of Two Markets*, 15 *Psychol. Sci.* 787, 790–91 (2004).

2009]

Making Good on Good Intentions

1939

ber of stop signs, posted speed limits, and other road regulations train drivers to drive by the signs rather than develop their own good judgment.¹⁶⁸ The theory, again, is that rules thwart the internalization process, which itself increases reliance on rules.¹⁶⁹

Increased legal coercion also works at odds with the internalization of desirable norms by signaling that other people would not otherwise adhere to those norms. This message is a negative influence,¹⁷⁰ to the extent that it suggests that the underlying norm does not reflect people's actual commitments, but rather standards that people would not follow unless the law imposed them. This phenomenon could help explain why, in one study, showing a sexual harassment policy to men had the unexpected effect of activating, rather than suppressing, gender stereotypes.¹⁷¹

The point is not that external norms must be avoided because they necessarily undermine the internalization of norms.¹⁷² External norms cannot be avoided and are necessary to give content to the values and expectations we seek people to internalize. Internalization, however, is not an inevitable consequence of legal coercion and may even be stifled by it. Excessively controlling, intrusive, or alienating contexts displaces people's sense of their own moral compass. Excessive control also leads people to engage in defensive or self-protective processes, disown responsibility for negative outcomes, and blame others.¹⁷³

These negative reactions to coercion and threat are damaging specifically to people's ability to internalize the nondiscrimination norms necessary to combat implicit bias. Researchers have been able to distinguish between people who are motivated primarily to avoid issues of discrimination, and those motivated primarily to

¹⁶⁸ See John Staddon, *Distracting Miss Daisy*, 302 *The Atlantic* 102 (July/August 2008).

¹⁶⁹ The failure to recognize the tendency for incentives to increase the need for incentives when norms are not internalized is the major weakness of an account of social norms that rests on incentives alone. For such an account, see Scott, *supra* note 134.

¹⁷⁰ McAdams, *supra* note 151, at 3–24.

¹⁷¹ See Justine Eatenson Tinkler et al., *Can Legal Interventions Change Beliefs? The Effect of Exposure to Sexual Harassment Policy on Men's Gender Beliefs*, 70 *Soc. Psychol. Q.* 480, 491 (2007).

¹⁷² See Deci & Ryan, *The "What" and "Why" of Goal Pursuits*, *supra* note 20, at 236–38.

¹⁷³ *Id.* at 229; Ryan & Deci, *Self-Determination Theory*, *supra* note 143, at 73.

avoid prejudice itself.¹⁷⁴ Individuals motivated primarily by a desire to escape the disapproval of others, stay out of trouble, and avoid conflict are externally motivated. In contrast, people who seek to avoid prejudice have integrated the goal of nondiscrimination into their other values; they seek to enact these values, not simply avoid the negative reactions of the law, employers, family, or friends.

Studies show that individuals motivated more by a desire to avoid dispute than to avoid prejudice tend to stay away from situations in which they might interact with blacks, especially when the situation is unscripted or uncertain.¹⁷⁵ These individuals are more likely to imagine unpleasant interactions, and thus more likely to avoid them.¹⁷⁶ In contrast, those with a strong internal motivation to avoid prejudice itself are more likely to imagine positive interactions with blacks, even when the scripts or expectations are less clear, and thus more likely to engage others not like themselves.¹⁷⁷

¹⁷⁴ To measure the motivation to avoid prejudice, the Motivation to Control Prejudiced Reactions scale uses a questionnaire that asks participants to indicate on a five-point scale his or her level of agreement with such statements as, "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 own viewpoint," or her level of disagreement with such statements as "I think that it is important to speak one's mind rather than to worry about offending someone." The motivation to avoid prejudice is measured by the individual's extent of agreement with such statements as, "I feel guilty when I have a negative thought or feeling about a Black person," or "I get angry with myself when I have a thought or feeling that might be considered prejudiced." Bridget C. Dunton & Russell H. Fazio, *An Individual Difference Measure of Motivation to Control Prejudiced Reactions*, 23 *Personality & Soc. Psychol. Bull.* 316, 319 (1997).

¹⁷⁵ Tamara Towles-Schwen & Russell H. Fazio, *Choosing Social Situations: The Relation Between Automatically Activated Racial Attitudes and Anticipated Comfort Interacting With African Americans*, 29 *Personality & Soc. Psychol. Bull.* 170, 178 (2003). Possible influences of the different motivations are suggested by the fact that those with high desire to avoid dispute reported less frequent interactions in school with blacks, admitted that their exposure to blacks during childhood was primarily through television and the media, and rated their parents as more prejudiced. In contrast, those with a high concern to avoid prejudice reported parental emphasis on egalitarianism, high scores on tests of egalitarian values, and relatively positive interactions with blacks during elementary and middle school. *Id.* at 172.

¹⁷⁶ *Id.* at 179–80.

¹⁷⁷ *Id.* Another difference is that people with a strong concern for avoiding prejudice are more likely to "overcorrect" in evaluating blacks as a group, while those with a strong concern for avoiding conflict are more likely to overcorrect for judgments about individual blacks, which researchers attribute to the possibility that making a situation more personal evokes a greater likelihood that a dispute will arise. See Michael A. Olson & Russell H. Fazio, *Trait Inferences as a Function of Automatically Activated Racial Attitudes and Motivation to Control Prejudiced Reactions*, 26 *Basic*

2009]

Making Good on Good Intentions

1941

This difference suggests that those individuals with more internalized nondiscrimination values—whose interactions are already the most positive—are the ones most likely to put themselves in situations where their prejudiced attitudes will be (still further) reduced.¹⁷⁸

Research on gender attitudes also has distinguished between internal and external motivation and found similar patterns. In one set of studies, test subjects whose desire to avoid sexism was based on external pressure rather than internalized norms rated sexist jokes negatively only when the circumstances risked disapproval by others. When the circumstances supported a sexist response (for example, the joke teller appearing to like the joke) and when the response was private, only test subjects whose non-sexist norms were personally important to them showed a reduced level of sexist response to sexist jokes. The other test subjects did not.¹⁷⁹

The premise of most of the legal reforms designed to make it easier for plaintiffs to win their cases is that people can try harder if the law leans on them more heavily. The motivation research discussed above, however, suggests that law alone will not activate the responses need to combat subtle, discriminatory behaviors, and may even undermine them. Part IV explores these principles further, in the context of research about what factors seem to matter in reducing these kinds of behaviors.

IV. STRATEGIES TO REDUCE IMPLICIT BIAS

Although social psychologists frequently note the difficulties of controlling stereotypes that are by definition unconscious and automatic,¹⁸⁰ most experts conclude that stereotypes are not per-

& Applied Soc. Psychol. 1, 9–10 (2004). On the overcorrection phenomenon, see supra text accompanying notes 78–87.

¹⁷⁸ See E. Ashby Plant, Responses to Interracial Interactions Over Time, 30 Personality & Soc. Psychol. Bull. 1458, 1469–70 (2004) (people who are internally motivated to avoid prejudice have more positive relationships with people of other races, and thus have better expectancies and less anxiety about future contact). Women tend to test higher than men for internally motivated avoidance of prejudice. See Plant & Devine, supra note 149.

¹⁷⁹ See Suzanne C. Klonis et al., Internal and External Motivation to Respond Without Sexism, 31 Personality & Soc. Psychol. Bull. 1237, 1246 (2005).

¹⁸⁰ See, e.g., Bargh, The Cognitive Monster, supra note 19, at 362, 376, 378 (controlling the influence of unconscious stereotypes is “almost unimaginable”); William T.

1942

Virginia Law Review

[Vol. 95:1893]

manent, but rather alterable.¹⁸¹ The two variables most often noted are (1) effort,¹⁸² and (2) situational context.¹⁸³ Both effort and situation matter,¹⁸⁴ and motivation plays a key role in determining how they do so, as this Part explains.

A. Effort

Research about the role of effort in reducing stereotypes appears to conflict. Some studies show that the effort to stop stereotyping might actually increase it, while others conclude that effort can have a beneficial effect on reducing stereotyped thinking.

Bielby, *Minimizing Workplace Gender and Racial Bias*, 29 *Contemp. Soc.* 120, 122 (2000) (“The task is not to eliminate ‘stereotypical thinking’ (it can’t be done), but rather to minimize its *impact* on personnel decisions.”); Timothy D. Wilson et al., *Mental Contamination and the Debiasing Problem*, in *Heuristics and Biases: The Psychology of Intuitive Judgment* 185, 190 (Thomas Gilovich et al. eds., 2002) (expressing a lack of optimism about people detecting and correcting for bias in everyday life).

¹⁸¹ See, e.g., Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 *Personality & Soc. Psychol. Rev.* 242 (2002). For a review of some of the literature, see Patricia G. Devine & Margo J. Monteith, *Automaticity and Control in Stereotyping*, in *Dual-Process Theories*, supra note 19, at 339–42.

¹⁸² See, e.g., Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 *J. Personality & Soc. Psychol.* 5, 6–7 (1989); Daniel T. Gilbert et al., *Unbelieving the Unbelievable: Some Problems in the Rejection of False Information*, 59 *J. Personality & Soc. Psychol.* 601, 611 (1990); Mitchell, supra note 153, at 687–88.

¹⁸³ Dual-process theories of stereotyping presuppose a pre-conscious (automatic) and a conscious (controlled) process, but sometimes the boundary between the two is softened by a “counterintuitive” notion that intent, will, and control “can operate without awareness.” See Gordon B. Moskowitz et al., *The History of Dual-Process Notions, and the Future of Preconscious Control*, in *Dual-Process Theories*, supra note 19, at 12, 33. In this regard, one set of researchers makes a helpful distinction between “deliberative and theory-driven” correction (or “debiasing”) and “rapid and nonconscious” correction (or “implicit adjustment”). See Wilson et al., supra note 180, at 185, 188–89.

¹⁸⁴ Mitchell, supra note 153, at 695–96 (identifying a broad continuum of processes through which attitudes can change, ranging from deliberate, conscious effort, to detectable (if usually unnoticed) thoughts and habits, to unconscious thinking styles and situation-induced goals). This observation is consistent with a variety of stereotype control strategies, from suppressing them once they arise, to actively seeking out information beyond the stereotype, overriding a stereotypic response with a clearly identifiable egalitarian response, or correcting a response in a direction opposite from the presumed influence of the stereotype. Studies demonstrating these various strategies are cited in Margo J. Monteith et al., *Putting the Brakes on Prejudice: On the Development and Operation of Cues for Control*, 83 *J. Personality & Soc. Psychol.* 1029, 1030 (2002).

2009]

Making Good on Good Intentions

1943

The research showing that effort increases rather than decreases stereotyping suggests that the concentration on ignoring a characteristic upon which a stereotype is based heightens the salience of the characteristic, thereby making the stereotype harder to control. In addition, the effort to appear unprejudiced to others can heighten anxiety about the characteristic, thereby triggering the automatic, nonverbal behaviors that this anxiety produces. Alone or together, these cognitive and attitudinal factors can lead to what is called a suppression “rebound effect,”¹⁸⁵ in which even if effort suppresses stereotypes in the first instance, stereotypes return with greater force when the pressure is relaxed.¹⁸⁶

One set of experimental studies supporting the existence of the rebound effect found that test subjects who were warned against using stereotypes were able to do so in specific exercises, but then when given no such warning in subsequent exercises, they exhibited more stereotypes than the control group.¹⁸⁷ When shown out of the room “after” one set of suppression exercises, test subjects who had suppressed stereotypes during the exercise chose seats further away from the black man whose picture had been used in the exercises.¹⁸⁸ The rebound effect can be subtle. In one experimental study, whites who tried hardest to appear color-blind were less likely to use race as a descriptor when describing other individuals to black partners, but this reluctance impaired communication and performance; these whites also made less eye contact with and appeared less friendly to their black partners.¹⁸⁹

¹⁸⁵ See C. Neil Macrae et al., *Out of Mind but Back in Sight: Stereotypes on the Rebound*, 67 *J. Personality & Soc. Psychol.* 808, 808–09 (1994).

¹⁸⁶ See Kerry Kawakami et al., *Just Say No (to Stereotyping): Effect of Training in the Negation of Stereotypic Associations on Stereotype Activation*, 78 *J. Personality & Soc. Psychol.* 871 (2000); Macrae, *supra* note 185, at 814.

¹⁸⁷ Macrae, *supra* note 185, at 811.

¹⁸⁸ *Id.* at 812. Another related effect is that the effort to produce one stereotype may alter the stereotype rather than eliminate it. See Sei Jin Do et al., *Sneaking in Through the Back Door: How Category-Based Stereotype Suppression Leads to Rebound in Feature-Based Effects*, 44 *J. Experimental Soc. Psychol.* 833, 833–34 (2008). For example, pressure to reduce race-based stereotyping might lessen stereotyping of blacks whose physical and social profiles are more similar to whites, while increasing it with respect to blacks with more Afrocentric features, or weaker resumes.

¹⁸⁹ See Michael I. Norton et al., *Color Blindness and Interracial Interaction: Playing the Political Correctness Game*, 17 *Psychol. Sci.* 949, 952 (2006).

Experiments testing the effect of trying to suppress sex-based stereotypes have shown comparable results. In one study, male and female subjects were asked to complete a series of sentence fragments, such as “Women who go out with a lot of men are”¹⁹⁰ In the study, subjects who were given the instruction to try not to be sexist in their responses gave fewer sexist responses (for example, “sluts”) and more “egalitarian” responses (for example, “popular”) when they were given time to think. When immediate responses were required, however, they gave more sexist responses when instructed not to be sexist than when given no instruction at all.¹⁹¹

In contrast to this research are studies demonstrating that attention and effort can have a positive effect on altering stereotypes.¹⁹² For example, experimental studies have shown that blatant priming tips people off that they may be subject to a potentially biasing influence, thereby triggering a successful effort to avoid stereotyped judgments.¹⁹³ Rote training exercises associating nonstereotypic traits with men and women also has been shown to reduce stereotyping, although the effect is not shown if the trainees make an effort to resist the training.¹⁹⁴ In a simulated job interview setting, interviewers who were given special encouragement to form accurate impressions exhibited less bias than those who were given no such encouragement; they also conducted longer interviews and asked

¹⁹⁰ Daniel M. Wegner, *Ironic Processes of Mental Control*, 101 *Psychol. Rev.* 34, 46–47 (1994). The study’s questions were drawn from the Attitudes Toward Women Scale developed by Spence and Helmreich. Janet T. Spence & Robert Helmreich, *The Attitudes Toward Women Scale: An Objective Instrument to Measure Attitudes Toward the Rights and Roles of Women in Contemporary Society*, JSAS Catalog of Selected Documents in Psychology (1972).

¹⁹¹ See Wegner, *supra* note 190, at 46–47. Wegner theorizes that this effect is caused by two different processes at work: an “operating” process that searches for mental content consistent with the desired state, and a “monitoring” process that seeks out, and thereby reinforces, inconsistent information. *Id.* at 34. Wegner calls this explanation “ironic process theory.” *Id.*

¹⁹² Blair, *supra* note 181, at 252.

¹⁹³ See Leonard S. Newman & James S. Uleman, *Assimilation and Contrast Effects in Spontaneous Trait Inference*, 16 *Personality & Soc. Psychol. Bull.* 224, 236 (1990); Fritz Strack et al., *Awareness of the Influence as a Determinant of Assimilation Versus Contrast*, 23 *Eur. J. Soc. Psychol.* 53, 59 (1993).

¹⁹⁴ See, e.g., Kerry Kawakami et al., *Kicking the Habit: Effects of Nonstereotypic Association Training and Correction Processes on Hiring Decisions*, 41 *J. Experimental Soc. Psychol.* 68, 73–74 (2005) (rote training in associating nonstereotypic traits with men and women can reduce stereotyping, unless trainees make an effort to resist training).

2009]

Making Good on Good Intentions

1945

better questions, thus gathering more and better information.¹⁹⁵ In other studies, telling people that they were going to be held accountable for the accuracy of a decision, and to whom, had a positive effect on non-stereotyped decisionmaking.¹⁹⁶ Likewise, job performance exercises instructing people to recall and write out both positive and negative behaviors of the person before reaching a performance rating appeared to reduce the impact of both race and gender bias.¹⁹⁷ In other research, asking people to write narrative essays about individuals whose perspective they were asked to assume led them to express more positive evaluations of those individuals and make fewer stereotypical comments, as compared to subjects given no special instructions, or those who were told simply to try to avoid stereotypic preconceptions.¹⁹⁸ These studies in various ways lend support to various “dual-process” theories of stereotyping positing that people, with effort or time for reflection, can overcome their first, stereotyped response to a stimulus through more thoughtful and deliberative cognitive processing.¹⁹⁹

¹⁹⁵ Neberg, *The Goal of Forming Accurate Impressions*, supra note 55, at 378–79; see also Susan T. Fiske & Steven L. Neberg, *A Continuum of Impression Formation, From Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation*, in 23 *Advances in Experimental Social Psychology* 1, 49 (Mark P. Zanna ed., 1990); Neberg, *Expectancy-Confirmation Processes*, supra note 45, at 103, 111–14. On the importance of strategies that encourage people to obtain more information in order to form a more accurate assessment, see Monteith et al., supra note 184, at 1030.

¹⁹⁶ See John Pennington & Barry R. Schlenker, *Accountability for Consequential Decisions: Justifying Ethical Judgments to Audiences*, 25 *Personality & Soc. Psychol. Bull.* 1067 (1999); Philip E. Tetlock & Jennifer S. Lerner, *The Social Contingency Model: Identifying Empirical and Normative Boundary Conditions on the Error-and-Bias Portrait of Human Nature*, in *Dual-Process Theories*, supra note 19, at 578; see also Leigh Ann Vaughn et al., *When Two Wrongs Can Make a Right: Regulatory Nonfit, Bias, and Correction of Judgments*, 42 *J. Experimental Soc. Psychol.* 654, 659 (2006) (accuracy instructions improve correction of bias).

¹⁹⁷ See Boris B. Baltes et al., *Does a Structured Free Recall Intervention Reduce the Effect of Stereotypes on Performance Ratings and by What Cognitive Mechanism?*, 92 *J. Applied Psychol.* 151, 157, 159, 161 (2007) (discussing race bias); Cara C. Bauer & Boris B. Baltes, *Reducing the Effects of Gender Stereotypes on Performance Evaluations*, 47 *Sex Roles* 465, 468 (2002) (discussing gender bias).

¹⁹⁸ Adam D. Galinsky & Gordon B. Moskowitz, *Perspective-Taking: Decreasing Stereotype Expression, Stereotype Accessibility, and In-Group Favoritism*, 78 *J. Personality & Soc. Psychol.* 708, 720 (2000).

¹⁹⁹ For articles on dual-process theories of overcoming prejudice, see *Dual-Process Theories in Social Psychology*, supra note 183. The two stages are sometimes distinguished as System I and System II thinking. See Jolls & Sunstein, supra note 94, at

Viewing this research through a motivation lens helps to make sense of the apparently conflicting findings about the role of effort. To the extent that individual effort depends upon external pressure, motivation principles would predict that behavior will conform to the desired norms only as long as that pressure exists, and that once the pressure lets up, autonomy will be reasserted and behavior will revert to prior habits. When the pressure is great enough or conflicts with existing beliefs and attitudes, these principles would also predict that attitudes might regress, as the individual pushes back in reaction to the resented control.

Likewise, it should be no surprise that individual effort that reflects, or enhances, an internalized commitment to nondiscrimination norms or overlaps with other goals, such as accurate outcomes or responsibility or accountability toward norms by respected others, would strengthen or reinforce those norms. Effort that produces information or encourages a more open or empathetic perspective would be expected to support the internalization of nondiscrimination norms, and thus to have a more positive and more lasting effect than pressure that is not supported by actual beliefs or attitudes.²⁰⁰ Even not-so-subtle priming would be expected to produce better results than more coercive influence, insofar as it allows individuals to identify and assert their own, conscious standards, rather than react to the imposed standards of others.²⁰¹

Some experimental research also suggests that how the effort to avoid stereotypes is evoked can affect the motivation of the effort, and thus the difference that it makes. For example, efforts that are promoted with positive signals rather than negative ones seem to

975; Daniel Kahneman & Shane Frederick, Representativeness Revisited: Attribute Substitution in Intuitive Judgment, *in* *Heuristics and Biases: The Psychology of Intuitive Judgment*, supra note 180, at 49, 50. Gregory Mitchell characterizes the difference as “initial thoughts” and “second thoughts.” Mitchell, supra note 153, at 687–88. For another distinction, see supra note 183.

²⁰⁰ Galinsky & Moskowitz, supra note 198, at 709, 720 (describing how perspective-taking motivates people by increasing the overlap between the self and the target of the perspective-taking).

²⁰¹ On the connection between the internalization of motivation and its effect on efforts to reduce bias, see Crandall et al., supra note 155, at 372–73; M. Michelle Peruche & E. Ashby Plant, Racial Bias in Perceptions of Athleticism: The Role of Motivation in the Elimination of Bias, 24 *Soc. Cognition* 438, 449–50 (2006).

2009]

Making Good on Good Intentions

1947

have a better effect.²⁰² Again, this contrast makes sense once differences in motivation are taken into account. Positive signals affirm and thus strengthen people's sense of identity with the desired norms, and thus help them to internalize those norms. Negative signals put people at odds with the authority, make them defensive about their failure to satisfy the authority's norms, and thus alienates them from those norms.

B. Exposure and Contact

Exposure to others, like effort, can have either a positive or a negative impact on intergroup attitudes. People can come in contact with others at various levels of intensity, ranging from merely observing counterstereotyped images, being in the presence of those presumed to have certain beliefs, or working closely with people from other groups. At each of these levels, intergroup exposure can threaten, or strengthen, feelings of autonomy, competence, and relatedness, and thus thereby affect people's motivation to avoid discriminatory behavior.

1. Counterstereotypes

Even intergroup exposure that is remote and impersonal can challenge and help alter stereotypes or attitudes that people hold about members of other groups. For example, priming exercises that pair positive words with black faces and negative words with white faces can reduce test subjects' scores on implicit bias tests.²⁰³ So can images of admired blacks like Tiger Woods, Denzel Washington, and Michael Jordan, disliked whites like Jeffrey Dahmer and Timothy McVeigh,²⁰⁴ and portrayals of blacks in positive set-

²⁰² The research is summarized in Gaertner et al., *supra* note 23, at 391–92. See also Gaertner & Dovidio, *Understanding and Addressing Contemporary Racism*, *supra* note 8, at 615, 633 (promoting positive thoughts works better than avoiding negative ones to improve interracial attitudes). For a more specific application of this point, see *infra* text accompanying notes 287–292.

²⁰³ See Michael A. Olson & Russell H. Fazio, *Reducing Automatically Activated Racial Prejudice through Implicit Evaluative Conditioning*, 32 *Personality & Soc. Psychol. Bull.* 421, 429 (2006). The effects persisted throughout a two-day separation between the conditioning procedure and the administration of the IAT. *Id.*

²⁰⁴ See Nilanjana Dasgupta & Anthony G. Greenwald, *On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and*

tings such as family barbeques or church, as compared to negative settings like a dilapidated street corner.²⁰⁵

Experiments that expose test subjects to members of minority groups in roles of authority also undermine customary stereotypes in some research settings. For example, a number of studies have shown that white participants tend to exhibit less automatic stereotype activation on implicit bias tests in the presence of an experimenter who is black rather than white.²⁰⁶ Even showing blacks in a neutral setting may dilute stereotypes.²⁰⁷ Likewise, exposure to female leadership can reduce gender stereotyping. For example, among women students who started college with comparable scores on automatic bias tests, those who attended women's colleges where they had frequent contact with women faculty showed less automatic bias after one year than those who attended coeducational institutions where the contact with women leaders was relatively less frequent.²⁰⁸ In another study, asking test subjects to

Disliked Individuals, 81 *J. Personality & Soc. Psychol.* 800, 806–07 (2001). The effect persisted when tested twenty-four hours later. *Id.* at 807.

²⁰⁵ See Bernd Wittenbrink et al., Spontaneous Prejudice in Context: Variability in Automatically Activated Attitudes, 81 *J. Personality & Soc. Psychol.* 815, 817, 823–24 (2001).

²⁰⁶ See Russell H. Fazio et al., Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?, 69 *J. Personality & Soc. Psychol.* 1013 (1995); Brian S. Lowery et al., Social Influence Effects on Automatic Racial Prejudice, 81 *J. Personality & Soc. Psychol.* 842, 845 (2001). The Lowery study, however, showed no reduction in automatic prejudice by Asian-American test-takers as measured by the IAT, except in a subsequent experiment in which some participants were instructed to avoid prejudice (or not) in the presence of the black experimenter. *Id.* at 847–49.

²⁰⁷ In one experiment, negative black stereotypes diminished after participants were exposed to a videotape of an interview with a black person for an average of ten minutes or more. See Ziva Kunda et al., The Dynamic Time Course of Stereotype Activation: Activation, Dissipation, and Resurrection, 82 *J. Personality & Soc. Psychol.* 283, 295 (2002). In this study, the staging of even a relatively trivial disagreement with the interviewee resurrected the stereotype, as measured in a test for implicit bias, but the activation of the stereotype did not appear to affect the participants' evaluation of the interviewee; in other words, stereotypes were activated, but apparently suppressed. *Id.* at 295–96.

²⁰⁸ See Nilanjana Dasgupta & Shaki Asgari, Seeing is Believing: Exposure to Counterstereotypic Women Leaders and Its Effect on the Malleability of Automatic Gender Stereotyping, 40 *J. Experimental Soc. Psychol.* 642, 653–54 (2004); see also M. Elizabeth Tidball et al., Taking Women Seriously: Lessons and Legacies for Educating the Majority (1999) (finding a strong link between frequency of counterstereo-

2009]

Making Good on Good Intentions

1949

imagine strong women—what they are like, why they are strong, what they are capable of doing, and what kinds of hobbies and activities they enjoy—reduced their implicit gender bias test scores.²⁰⁹

As would be expected, the strength and duration of these effects are limited by the degree of exposure. When exposure is brief, so are its effects.²¹⁰ Frequency of contact and salience would be expected to improve the effectiveness of counterstereotypes.²¹¹ Thus, publicly visible portraits and statues are likely to have greater positive impact than single exposures. Likewise, when women and minorities are in high positions for a sustained period and with a record of success, their impact on stereotypes will likely be greater than when circumstances seem to suggest that their counterstereotypical role was a fluke, or when it appears to confirm the stereotype that women and minorities are not meant for positions of responsibility in the public realm.

Positive effects also may be limited by how broadly a countervailing influence is generalized to the group as a whole. As noted earlier, a feature of stereotypes is that they tend to repel disproving data; people often process information that conflicts with a stereotype as an exception to a robust rule of thumb, rather than a reason to modify the stereotype.²¹² Thus, Tiger Woods might elicit positive feelings about himself as a talented individual or perhaps increase positive attitudes about black athletes, but his example may not improve people's judgments about blacks in general. He is viewed as an outlier—someone who proves blacks *can* be admirable peo-

typic female role models on campus (for example, faculty, administrators, and peers) and the cultivation of students' commitment to counterstereotypic careers).

²⁰⁹ Irene V. Blair et al., *Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery*, 81 *J. Personality & Soc. Psychol.* 828, 837 (2001). For a review of other studies, see Blair, *The Malleability of Automatic Stereotypes and Prejudices*, *supra* note 181, at 248–49.

²¹⁰ Dasgupta & Greenwald, *supra* note 204, at 806–07 (improved effect of images of admired blacks and dislike whites on automatic stereotypes was significantly diminished twenty-four hours later).

²¹¹ On the concept of “construct-accessibility,” see E. Tory Higgins & Gillian King, *Accessibility of Social Constructs: Information-Processing Consequences of Individual and Contextual Variability*, *in* *Personality, Cognition, and Social Interaction* 69, 71 (Nancy Cantor & John F. Kihlstrom eds., 1981); Thomas K. Srull & Robert S. Wyer, Jr., *Category Accessibility and Social Perception: Some Implications for the Study of Person Memory and Interpersonal Judgments*, 38 *J. Personality & Soc. Psychol.* 841 (1980).

²¹² See *supra* text accompanying notes 46 and 55.

1950

Virginia Law Review

[Vol. 95:1893]

ple, but not that they usually *are*.²¹³ For generalization to occur, remote or passing exposure to counter-stereotypes may not be enough.

2. *Minimal Contact and Peer Influence*

Researchers have shown, in many different contexts, that the physical presence of other people can attenuate bias. In university settings, for example, students revealed less bias on an implicit bias test if they took the test in the physical presence of other students than if they took the test alone.²¹⁴ The presence of blacks seems particularly to motivate whites to be more attentive to race issues.²¹⁵ For example, students noticed more prejudice in a film clip when they were told that they were being watched by one or two blacks in the adjacent room.²¹⁶ In an experiment involving a simulated trial of a black defendant, whites who sat on racially mixed juries tended to discuss more case facts, were less likely to challenge the possibility of racism, and were more likely to vote for acquittal.²¹⁷

Some researchers explain this effect by a theory of “social tuning.” Social tuning theory posits that people generally prefer to have positive interactions with others, and that they bring their own attitudes in line with the presumed views of others in order to have those positive interactions.²¹⁸ Some researchers describe a

²¹³ In addition to the social psychology experiments demonstrating this phenomenon, there is some physiological support. See E.A. Phelps et al., Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation, 12 *J. Cognitive Neuroscience* 729, 733–34 (2000) (while the amygdalae of whites typically over-react to black faces in experimental settings, they do not do so with famous black faces). For citations to research on strategies to promote the generalization of counterstereotypes, see *infra* note 238.

²¹⁴ See Luigi Castelli & Silvia Tomelleri, Contextual Effects on Prejudiced Attitudes: When the Presence of Others Leads to More Egalitarian Responses, 44 *J. Experimental Soc. Psychol.* 679, 683–84 (2008).

²¹⁵ Robinson, *supra* note 8, at 1177.

²¹⁶ See Daisuke Akiba & Payneese Miller, The Expression of Cultural Sensitivity in the Presence of African Americans: An Analysis of Motives, 35 *Small Group Res.* 623, 637 (2004).

²¹⁷ See Samuel R. Sommers, On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations, 90 *J. Personality & Soc. Psychol.* 597, 603–06 (2006).

²¹⁸ See Lowery et al., *supra* note 206, at 843; Stacey Sinclair et al., Social Tuning of the Self: Consequences for the Self-Evaluations of Stereotype Targets, 89 *J. Personal-*

2009]

Making Good on Good Intentions

1951

more general tendency of people to assimilate to a perceived socially shared norm.²¹⁹ Consistent with both theories, it appears that the likeability of others enhances people's motivation to adapt to the attitudes they are thought to have.²²⁰ Likewise, having power over others appears to weaken that motivation; people with power show more implicit bias toward their subordinates than those interacting with someone who is their superior or co-equal.²²¹

People's attitudes about members of other groups are especially influenced by people they perceive to be their peers, whose views often appear to be more important than people's own personal experiences.²²² Like other situational factors, these influences can either reinforce bias, or reduce it.²²³

ity & Soc. Psychol. 160 (2005). Social tuning is exhibited even in simple exchanges, such as when people use social categorization cues (such as accents and clothing) to determine how to respond to a request for local directions. See Ellen A. Isaacs & Herbert H. Clark, References in Conversation Between Experts and Novices, 116 J. Experimental Psychol.: Gen. 26, 31 (1987).

²¹⁹ See Castelli & Tomelleri, *supra* note 214, at 684; see also Crandall et al., *supra* note 155, at 361; Gretchen B. Sechrist & Charles Stangor, Perceived Consensus Influences Intergroup Behavior and Stereotype Accessibility, 80 J. Personality & Soc. Psychol. 645, 651 (2001).

²²⁰ See Sinclair et al., Social Tuning of the Self, *supra* note 218. These attitudes can be either positive or negative ones. See Castelli & Tomelleri, *supra* note 214, at 684; see also Crandall et al., *supra* note 155, at 361; Sechrist & Stangor, *supra* note 219, at 651.

²²¹ Stacey Sinclair et al., Social Tuning of Automatic Racial Attitudes: The Role of Affiliative Motivation, 89 J. Personality & Soc. Psychol. 583, 584 (2005); see also Richard Y. Bourhis, Power, Gender, and Intergroup Discrimination: Some Minimal Group Experiments, *in* The Psychology of Prejudice, *supra* note 38, at 200 (making the same conclusion with respect to gender bias); Jennifer A. Richeson & Nalini Ambady, Effects of Situational Power on Automatic Racial Prejudice, 39 J. Experimental Soc. Psychol. 177 (2003) (noting that in experimental studies, people assigned superior roles tend to register higher on implicit race bias tests and exhibit more stereotyping behavior when evaluating others as compared to those who hold equal or subordinate status with those they are evaluating). The differential effects may be related to why interaction between individuals who function as equals is more successful in reducing prejudice than in hierarchical relationships. See *infra* text accompanying note 232; see also *supra*, text accompanying note 150 (people who need to dominate situations are more prone to biased evaluations of others).

²²² See, e.g., Bruno Lasker, Race Attitudes in Children 371 (1929); see also Muzafer Sherif & Carolyn W. Sherif, Groups in Harmony and Tension 94–95 (1953) (developing “group norm theory” to explain the pressures on individuals to conform to group norms); Sechrist & Stangor, *supra* note 219 at 645.

²²³ See Sherif & Sherif, *supra* note 222; see also Crandall et al., *supra* note 155, at 361 (concluding that social norms are strong predictors of expressed prejudice).

1952

Virginia Law Review

[Vol. 95:1893]

Research in university campus settings has demonstrated in particular the importance of peer influence on attitudes about race. In various experiments, students' attitudes about race were affected by exposure to the views of other students.²²⁴ A series of experiments involving students from the University of Maryland suggest that peer influence about race outweighs more rational or evidence-based factors. One experiment had students estimate (1) the percentage of black students who possessed 16 stereotypical traits and (2) how the same judgments would be made by the average University of Maryland student. One week later, some students were given feedback suggesting that other University of Maryland students shared their beliefs and others were given the opposite feedback. The students who were told that their peers had more favorable views of blacks than they had predicted themselves registered more favorable attitudes when retested. Those told that their peers had more negative views than they had predicted registered more negative stereotypes when tested again.²²⁵ In a follow-up experiment, students were exposed to supposedly "scientific information" about what percentage of blacks possessed each trait. In later testing, whether the "scientific information" mattered to their own views, or their estimates of other people's views, depended upon the consistency between peer attitudes and their own. When peer attitudes supported the students' original estimates, the data did not change them. When they did not, the data had more influence on students' subsequent views.²²⁶

The greater the identification with peers, the more influence peers have on beliefs and attitudes. For example, students in one study were more influenced by the attitudes of students at their own college than they were by the attitudes of students from a rival

²²⁴ See, e.g., Fletcher A. Blanchard et al., *Condemning and Condoning Racism: A Social Context Approach to Interracial Settings*, 79 *J. Applied Psychol.* 993, 995 (1994) (hearing another student condemn racism, or condone it); Fletcher A. Blanchard et al., *Reducing the Expression of Racial Prejudice*, 2 *Psychol. Sci.* 101, 101-03 (1991) (hearing another student express their views about a campus incident with potential race overtones); Sechrist & Stangor, *supra* note 220, at 649-51 (being told of the racial views of other students).

²²⁵ Charles Stangor et al., *Changing Racial Beliefs by Providing Consensus Information*, 27 *Personality & Soc. Psychol. Bull.* 486, 489-90 (2001).

²²⁶ *Id.* at 492-93.

2009]

Making Good on Good Intentions

1953

college.²²⁷ Simply knowing of friendships between members of one's own group and those of another, however, may be enough to improve a person's attitude toward that other group.²²⁸ Notably, the University of Maryland research indicates that it is easier to influence attitudes in a positive direction than in a negative one.²²⁹

Learning that one's attitudes are shared by others has the potential to affect behavior as well as measured attitudes. In one of the University of Maryland studies, students indicated their attitudes about blacks on a computerized test, and then were given (false) feedback about the extent to which their responses agreed with other University of Maryland students. They were then (on a false pretense) led to a room and told to take a seat where a black female was already seated. Students who tested high for prejudice toward blacks and who were told that a high percentage of Maryland students agreed with them sat further away from the black female than students who had been told that only a minority of others agreed with them.²³⁰

3. *Working Relationships and the "Contact Hypothesis"*

Moving beyond visual images, mere physical presence, and peer influence, research experiments demonstrate that interaction between members of different groups in the workplace are, under the right circumstances, the most important factor in reducing group-based stereotypes in the workplace. In 1954, Gordon Allport identified a number of factors that appeared to determine whether group contact generated more positive attitudes toward that group. The factors set forth in Allport's "contact hypothesis" included equal status between the groups, common goals, the interdependence of the groups, and the positive support of authorities, laws, or custom.²³¹

²²⁷ Id. at 491–92.

²²⁸ See Stephen C. Wright et al., *The Extended Contact Effect: Knowledge of Cross-Group Friendships and Prejudice*, 73 *J. Personality & Soc. Psychol.* 73, 74 (1997).

²²⁹ Stangor et al., *supra* note 225, at 493.

²³⁰ Sechrist & Stangor, *supra* note 219, at 649.

²³¹ Dovidio et al., *supra* note 145, at 147; see Allport, *supra* note 38, at 281. Much research has affirmed the basic principles of the contact hypothesis. For reviews and updates, see *On the Nature of Prejudice: Fifty Years After Allport* 8–9 (John F. Dovidio et al. eds., 2005); John F. Dovidio et al., *Intergroup Bias: Status, Differentiation, and a Common In-Group Identity*, 75 *J. Personality & Soc. Psychol.* 109, 109–10

Follow-up research has confirmed the importance of Allport's four key conditions and refined the basic findings. Equal status within the situation matters; even when the people differ in status outside the interaction, having equal status within the interaction has a positive effect.²³² The existence of common goals also matters, as long as the effort toward achieving them is structured to be cooperative rather than competitive.²³³ And explicit support from authorities is highly significant; without it, the positive effects of equal status and common goals may be negated.²³⁴ Indeed, without all of these conditions working together, workplace interactions may reinforce group stereotyping rather than weaken it.²³⁵

One explanation for the positive impact of contact on intergroup bias is that it reduces the salience of race and sex.²³⁶ The underlying mechanism appears to be both cognitive and motivational. Cognitively, intergroup contact decategorizes group boundaries, so that group members view themselves and others more as either individuals, or part of one larger group—and less as members of separate, competing groups. Motivationally, forming a “common group identity” triggers the same positive attitudes that characterize in-group bias, therefore leading people to see each other more generously.²³⁷ This effect is greatest, some researchers have noted, if sufficient salience of the original groups is preserved, so that changed attitudes toward the individuals with whom people are in contact

(1998); Thomas F. Pettigrew & Linda R. Tropp, A Meta-Analytic Test of Intergroup Contact Theory, 90 *J. Personality & Soc. Psychol.* 751, 751 (2006).

²³² Thomas F. Pettigrew & Linda R. Tropp, Allport's Intergroup Contact Hypothesis: Its History and Influence, *in* *On the Nature of Prejudice*, supra note 231, at 265.

²³³ Id.; David W. Johnson & Roger T. Johnson, The Three Cs of Reducing Prejudice and Discrimination, *in* *Reducing Prejudice and Discrimination* 239, 249 (Stuart Oskamp ed., 2000) (concluding a meta-analysis of 180 studies shows strong effect of cooperative experiences over competitive ones in reducing prejudice).

²³⁴ Johnson & Johnson, supra note 233, at 249.

²³⁵ See, e.g., Cecilia L. Ridgeway & Lynn Smith-Lovin, The Gender System and Interaction, 25 *Ann. Rev. Soc.* 191, 209 (1999) (noting male-female interactions that are status-ordered “continually refresh[]” gender status beliefs).

²³⁶ See Samuel L. Gaertner et al., How Does Cooperation Reduce Intergroup Bias?, 59 *J. Personality & Soc. Psychol.* 692 (1990).

²³⁷ See generally Gaertner & Dovidio, Reducing Intergroup Bias: The Common In-group Identity Model, supra note 51.

2009]

Making Good on Good Intentions

1955

are generalized beyond the conditions of the interaction itself, to other members of the outside group.²³⁸

Intergroup contact also breaks down boundaries between groups and reduces people's anxiety about each other,²³⁹ motivating them to disclose things to each other, and even establish friendships.²⁴⁰ These ties promote empathy, and reduce the sense of threat and alienation that otherwise leads them to treat others as individuals rather than as members of alien groups, and thus have more positive attitudes toward one another.²⁴¹ This openness may help to explain, more generally, why diversity within a person's networks causes people to be less fixed in their own views, and more careful about assessing new information and attitudes.²⁴²

This research, taken together, reveals that reducing discriminatory attitudes entails a combination of cognitive and motivational factors. Effort can make a difference, if it is motivated positively,

²³⁸ See Rupert Brown et al., *Changing Attitudes Through Intergroup Contact: The Effects of Group Membership Salience*, 29 *Eur. J. Soc. Psychol.* 741 (1999); Matthew J. Hornsey & Michael A. Hogg, *Assimilation and Diversity: An Integrative Model of Subgroup Relations*, 4 *Personality & Soc. Psychol. Rev.* 143, 148 (2000); Alberto Voci & Miles Hewstone, *Intergroup Contact and Prejudice Toward Immigrants in Italy: The Mediation Role of Anxiety and the Moderational Role of Group Salience*, 6 *Group Processes & Intergroup Rel.* 37, 38–39 (2003); see also Christopher Wolsko et al., *Framing Interethnic Ideology: Effects of Multicultural and Color-Blind Perspectives on Judgments of Groups and Individuals*, 78 *J. Personality & Soc. Psychol.* 635, 648–49 (2000) (describing experiments suggesting that multi-cultural framework led to stereotypes that were stronger than those generated by color-blind perspective, but also more accurate and more attentive to individuating information).

²³⁹ Alberto Voci & Miles Hewstone, *supra* note 238, at 37–38.

²⁴⁰ See Nurcan Ensari & Norman Miller, *The Out-Group Must Not Be So Bad After All: The Effects of Disclosure, Typicality, and Salience on Intergroup Bias*, 83 *J. Personality & Soc. Psychol.* 313, 325 (2002); Thomas F. Pettigrew, *Intergroup Contact Theory*, 49 *Ann. Rev. Psychol.* 65, 72 (1998). The importance of self-disclosure in improving explicit attitude toward members of other groups through empathy and trust is explored in Rhiannon N. Turner et al., *Reducing Explicit and Implicit Outgroup Prejudice Via Direct and Extended Contact: The Mediating Role of Self-Disclosure and Intergroup Anxiety*, 93 *J. Personality & Soc. Psychol.* 369, 383–84 (2007).

²⁴¹ See Thomas F. Pettigrew, *Generalized Intergroup Contact Effects on Prejudice*, 23 *Personality & Soc. Psychol. Bull.* 173, 174 (1997) (describing “deprovincialization”); Pettigrew & Tropp, *A Meta-Analytic Test of Intergroup Contact Theory*, *supra* note 231, at 766 (concluding based upon a meta-analysis that attitudes changed from intergroup contact generalize to other members of the outgroup and in other situations).

²⁴² See Lindsey Clark Levitan & Penny S. Visser, *The Impact of the Social Context on Resistance to Persuasion: Effortful Versus Effortless Responses to Counter-Attitudinal Information*, 44 *J. Experimental Soc. Psychol.* 640, 646 (2007).

1956

Virginia Law Review

[Vol. 95:1893]

to build upon accepted values and internal norms. Situational influences are important to that motivation, the content of the norms, and the likelihood that people will internalize those norms. Part V brings these findings to bear on strategies for reducing implicit discrimination.

V. THE IMPLICATIONS OF MOTIVATION ANALYSIS FOR LEGAL AND STRUCTURAL REFORM

Part IV explained how effort and situational factors that inform people of the positive views of others and stimulate interactions between members of different groups have the potential to alter the attitudes and beliefs that cause discriminatory behavior. This Part evaluates the reform proposals that were outlined briefly in Part II in light of the social science evidence accumulated in Parts III and IV. It concludes that clear, enforceable, and enforced standards prohibiting workplace discrimination are necessary to reducing it, but that legal rules that are perceived as unfair, or that are experienced as unnecessarily coercive, will undermine people's needs for autonomy, competence, and relatedness, and thereby diminish their internalization of nondiscrimination values. Strong legal standards can deter provable instances of discrimination and compensate victims, and they define desirable norms. The law, however, cannot reach the more hidden and ambiguous forms of discrimination, no matter how forcefully it tries to prohibit them. Instead of relying entirely on threats and coercion, which may even make discrimination worse, nondiscrimination strategies must also take account of people's need for a sense of autonomy, competence, and relatedness.

A. Legal Reform Proposals

Proposals designed to clarify that a successful employment discrimination case under Title VII does not require a showing of deliberate race or gender animus do not raise the concerns highlighted in this Article. Knowing that one is responsible for one's actions—even if one is not aware that those actions are discriminatory—challenges people to be more thoughtful and careful about their decisions. Accordingly, Title VII was drafted and has been interpreted to prohibit behaviors that discriminate “because of” race,

2009]

Making Good on Good Intentions

1957

gender, or other protected characteristic, not just those that are consciously discriminatory.²⁴³ By definition, implicit discrimination would be put beyond the law's reach if showing that one had an honest belief that one was acting in a nondiscriminatory fashion at the time in question, or had so acted in the past, was an adequate defense.²⁴⁴ Good intentions are not a "free pass" to discrimination and should never exonerate a defendant from liability for conduct that adversely treats someone because of a characteristic protected by Title VII.²⁴⁵

Several proposals, however, not only make discriminatory intent irrelevant to a Title VII case, but they also abandon the necessity for establishing a causal link between an employment decision and a victim's race, sex, or other protected characteristic.²⁴⁶ These proposals alter the basic liability standard so that even a person who has not treated someone differently because of their race or sex may be found to have done so. Rules that impose liability without guilty intent are one thing; rules that impose liability for discrimination that has not occurred are another, and are likely to stimulate the negative effects examined in this Article.

Ann McGinley's proposed mandatory presumption against the employer if the employer offers an explanation for an employment decision that turns out to be false²⁴⁷ has this potential risk. Discrimination is too complex a phenomenon to try to prove or disprove through conclusive presumptions. Decisionmaking in the employment context is not monolithic, and supervisors frequently act for a number of reasons. As the law now provides, offering a nondiscriminatory justification that is later shown not to have been the "real reason" may support an inference of discrimination, but it should not be a conclusive basis for assuming it.²⁴⁸ If proof of honesty is not enough to rebut discrimination,²⁴⁹ neither should proof of falsehood prove it. Legally mandating a finding of discrimina-

²⁴³ See supra Section II.A.

²⁴⁴ See id.

²⁴⁵ Cf. *Griggs v. Duke Power Co.*, 401 U.S. 424, 432 (1971) ("good intent or absence of discriminatory intent does not redeem" practices that violate Title VII).

²⁴⁶ See supra Section II.B.

²⁴⁷ McGinley, supra note 17, at 482.

²⁴⁸ For examples of possible fact situations, see supra text accompanying note 121.

²⁴⁹ See Krieger & Fiske, supra note 99, at 1034-38 (criticizing the "honest belief" rule).

tion on the basis of a lie is no more fair than requiring that a lie is not enough,²⁵⁰ or mandating a finding for the defendant if the plaintiff is caught in a lie. The appearance of neutrality is essential to the moral force of Title VII, which in turn is essential to people's internalization of its nondiscrimination norms. An approach that ignores the costs of false findings of discrimination fosters a perception of unfairness, and the anxieties of being found guilty while innocent under these rules may drive people, inadvertently, to act in more stereotyped ways.

Russell Robinson's proposal for dual-track liability to provide relief in close cases that an "outsider" would find compelling, even if an "insider" would not,²⁵¹ also seems unnecessarily provocative, as well as impossible to implement even-handedly. The proposal highlights the important fact that insiders and outsiders see bias differently. Like proposed rules that alter burden of proof rules to eliminate the causation requirement, however, the Robinson proposal would short-circuit the question of whether a person was treated differently in the workplace "because of" a prohibited characteristic, and instead would assume a connection between a person's identity status and how they were treated in a particular case under a general theory of the prevalence of discrimination. It is one thing to say that discrimination should be found even if conscious intent was lacking. It is another to say that discrimination should be found because it often occurs but will not be provable without a rule that presumes it.²⁵² Proposals that convert ambiguity into certainty associate Title VII with a conspicuous rule of favoritism and encourage defensive, stereotyped reactions. We need to get better at identifying discrimination, without assuming that it exists in the absence of actual proof.

Proposals designed to hold employers liable for workplace procedures and structures that could give rein to unconscious bias, even if not shown to do so, have similar problems. David Oppenheimer proposes to have courts second-guess all workplace deci-

²⁵⁰ At one time, proving the lie was not enough. See *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 511 (1993). This rule was reversed in *Reeves v. Sanderson Plumbing Products, Inc.*, 530 U.S. 133, 134 (2000), holding that an inference of discrimination could be made based on the lie, although the inference is not required.

²⁵¹ Robinson, *supra* note 8, at 1166–67.

²⁵² See Wax, *The Discriminating Mind*, *supra* note 12.

2009]

Making Good on Good Intentions

1959

sions involving women and minorities.²⁵³ Tristin Green does the same with respect to workplace procedures that may inadvertently contribute to ingroup favoritism and outgroup bias, or that do not protect members of disadvantaged groups from responding badly to perceived bias.²⁵⁴ In some workplaces with entrenched habits, procedures, and attitudes, greater surveillance may be beneficial, but existing Title VII law should be able to identify these cases and mandate remedial regimes that address the problems.²⁵⁵ Imposing surveillance regimes on all workplaces ignores the impact of excessive suspicion and micromanagement on the kind of trust, organizational esprit, and commitment to nondiscrimination norms.²⁵⁶ Employer liability for workplace practices that could hypothetically permit biased decisions, but have not been shown to do so, pushes people into defensive, resentful responses. If the goal is good decisionmaking relative to race, gender, and other protected characteristics, this is the wrong “nudge.”²⁵⁷

Rules imposing employer liability based on potential rather than actual discrimination also will tend to generate from employers more of the standard liability-based bullet-proofing measures, such as diversity training and various mechanical reporting requirements that, as discussed in the next Section,²⁵⁸ have not proved effective. These measures are motivated more by the goal of avoiding costly judgments than by reducing discrimination. Added pressure may lead employers to adopt the proposal of some scholars that all prospective employees be given the Implicit Association Test.²⁵⁹ These tests are easy enough to administer, but are unlikely

²⁵³ Oppenheimer, *supra* note 94, at 970.

²⁵⁴ Green, *supra* note 18, at 856–57, 899–903.

²⁵⁵ *Dukes et al. v. Wal-Mart Stores, Inc.*, 474 F.3d 1214 (9th Cir. 2007), *granting rehearing en banc* 556 F.3d 919 (9th Cir. 2009) (certifying a class of 1.5 million female employees of Wal-Mart who alleged systematic discrimination in hiring and promotion practices).

²⁵⁶ Cialdini, *supra* note 160, at 56–57; see *supra* text accompanying note 160.

²⁵⁷ See Richard H. Thaler & Cass R. Sunstein, *Nudge* (2008) (urging more thoughtfulness to how decisions are structured for people, to encourage them to make decisions that are best for themselves and others).

²⁵⁸ See *infra* Section V.B.

²⁵⁹ See Ian Ayres, *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination* 424–25 (2001) (advocating IAT scores as a criterion for hiring both governmental and nongovernmental actors); Kang & Banaji, *supra* note 5, at 1091. Some have even argued in favor of using the IAT test in the course of litigation,

1960

Virginia Law Review

[Vol. 95:1893]

to amount to any meaningful workplace change—since almost everyone registers bias on these tests—and may only increase anxiety and denial, which enhances bias.²⁶⁰

B. Structural Reform Proposals

A comprehensive study of the efficacy of diversity promotion efforts, using federal data over a thirty-year period, concludes that diversity training and diversity evaluations are the least effective in enhancing diversity. Meanwhile, mentoring programs and networking efforts show modest positive effects, and accountability mechanisms show the greatest improvement in organization diversity.²⁶¹ Based on both experimental studies and field research, researchers have concluded also that collaborative workplace cultures,²⁶² the reduction in status differentials within work groups,²⁶³ and job rotation programs²⁶⁴ all seem to reduce stereotyping as well as improve prospects for women and members of racial minorities in the workplace.

A focus on promoting people's good intentions makes sense of these research conclusions. For example, researchers believe that it is important for people to be aware of their biases.²⁶⁵ Yet, while many have urged education as the primary means of improving at-

as probative of whether the defendant discriminated. See Audrey J. Lee, Unconscious Bias Theory in Employment Discrimination Litigation, 40 *Harv. C.R.-C.L. L. Rev.* 481, 501–02 (2005).

²⁶⁰ For the same reason, the suggested use of required, “de-biasing” screensavers is equally unpromising. For the suggestion, see Jerry Kang, Trojan Horses of Race, 118 *Harv. L. Rev.* 1489, 1536–37 (2005). Voluntary IAT testing and screensavers are, of course, another matter.

²⁶¹ Alexandra Kalev et al., Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies, 71 *Am. Soc. Rev.* 589, 611 (2006).

²⁶² See, e.g., Samuel B. Bacharach et al., Diversity and Homophily at Work: Supportive Relations Among White and African-American Peers, 48 *Acad. Mgmt. J.* 619, 620 (2005).

²⁶³ Green & Kalev, *supra* note 34, at 1445–49.

²⁶⁴ *Id.* at 1450–52.

²⁶⁵ See Devine & Monteith, *supra* note 181, at 346; Fiske, *supra* note 44, at 364; Wilson & Brekke, *supra* note 180, at 119–20; see also Jack Glaser & Eric D. Knowles, Implicit Motivation to Control Prejudice, 44 *J. Experimental Soc. Psychol.* 164, 171 (2008) (noting that people's implicit belief that they are prejudiced moderates the effect of unconscious stereotypes).

2009]

Making Good on Good Intentions

1961

titudes,²⁶⁶ training programs have a poor success record.²⁶⁷ Why is this? Based on what we know about what motivates people to change their stereotyped attitudes about other people, a reasonable hypothesis is that training programs too often communicate to their intended audiences the message that individuals need to be taught not to discriminate because right now they do not know any better. Most people already know that it is wrong to be prejudiced; the problem is that, left to their own devices, they too often unwittingly engage in it.²⁶⁸ “Preaching about the evils of prejudice would likely be ineffective [to people who] already agree with the message and regard themselves to be nonprejudiced.”²⁶⁹ In the vocabulary of motivation, telling people that they would discriminate if they were allowed to do so, or unless they were taught not to do so, undermines their senses of autonomy, competence, relatedness, and basic goodness.²⁷⁰ Diversity training programs ignore this in-

²⁶⁶ See, e.g., Faye J. Crosby & Susan Clayton, *Affirmative Action: Psychological Contributions to Policy*, 1 *Analyses Soc. Issues & Pub. Pol’y* 71, 81 (2001).

²⁶⁷ Rebecca S. Bigler, *The Use of Multicultural Curricula and Materials to Counter Racism in Children*, 55 *J. Soc. Issues* 687, 699–701 (1999) (concluding that diversity training often heightens rather than reduces race bias in children, and that use of counterstereotypes within the curriculum and more attention to tasks requiring children to attend to similarities and differences among individuals of different races would be more effective); Susan Bisom-Rapp, *An Ounce of Prevention Is a Poor Substitute for a Pound of Cure: Confronting the Developing Jurisprudence of Education and Prevention in Employment Discrimination Law*, 22 *Berkeley J. Emp. & Lab. L.* 1, 4 (2001) (noting little evidence that education and training programs work and arguing that they may have a polarizing effect on employee attitudes); Kalev et al., *supra* note 261, at 593–94 (reviewing studies); see also James A. Banks, *Multicultural Education: Its Effects on Students’ Racial and Gender Role Attitudes*, in *Handbook of Research on Multicultural Education* 617, 624 (James A. Banks & Cherry A. McGee Banks eds., 1995) (concluding that the effects of multicultural education programs are “inconsistent”).

²⁶⁸ See Bonita London et al., *Studying Institutional Engagement: Utilizing Social Psychology Research Methodologies to Study Law Student Engagement*, 30 *Harv. J.L. & Gender* 389, 395 (2007) (noting that law students recognize the value of diversity, but they fail to successfully integrate in their own voluntary intergroup actions).

²⁶⁹ Gaertner & Dovidio, *Understanding and Addressing Contemporary Racism*, *supra* note 8, at 626.

²⁷⁰ See *supra* text accompanying notes 136–139; see also Laurie A. Rudman et al., “Unlearning” Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes, 81 *J. Personality & Soc. Psychol.* 856, 857 (2001) (noting that people “perceive a threat to their freedom of expression or [are] offended by the implication that they are prejudiced”).

1962

Virginia Law Review

[Vol. 95:1893]

sight, and thus are highly unpopular among workers.²⁷¹ Programs that employers want courts to be able to easily recognize as remedial in nature tend to be poorly designed to motivate people to do the self-examination necessary to reduce unconscious bias. Instead, they raise hackles and are often the butt of office jokes.²⁷² The resistance to these programs, itself, increases the likelihood that they will backfire.²⁷³

In contrast, collaborative work cultures that minimize status differentials address the conditions that feed stereotyped thinking. Working together allows people to get to know each other and motivates them to form accurate assessments of one another, not stereotyped ones.²⁷⁴ Collaboration puts people in situations in which they are more likely to share the personal information upon which common bonds can be formed, and are thus less likely to stereotype. People feel more valued and secure in such institutional cultures²⁷⁵ and thus better able to internally absorb workplace values as their own.

²⁷¹ They appear to be especially unpopular among male employees. See Deborah L. Kidder et al., *Backlash Toward Diversity Initiatives: Examining the Impact of Diversity Program Justification, Personal and Group Outcomes*, 15 *Int'l J. Conflict Mgmt.* 77, 93 (2004) (noting that women are more positive than men toward diversity-related programs); Sara Rynes & Benson Rosen, *A Field Survey of Factors Affecting the Adoption and Perceived Success of Diversity Training*, 48 *Personnel Psychol.* 247, 262 (1995) (reporting that men perceive diversity programs as less successful than women do).

²⁷² See Patricia G. Devine et al., *Breaking the Prejudice Habit: Progress and Obstacles*, in *Reducing Prejudice and Discrimination* 185, 193 (Stuart Oskamp ed., 2000) (recounting reports of senior Texaco executives' derision of their own diversity training programs).

²⁷³ See Kawakami et al., *Kicking the Habit*, supra note 194, at 74.

²⁷⁴ See Fiske & Neuberg, supra note 195, at 38, 46–49; Norman Miller, *Personalization and the Promise of Contact Theory*, 58 *J. Soc. Issues* 387, 391 (2002). One refinement on this general principle is that one-on-one competition appears to facilitate individuating impressions of opponents, while group-on-group competition fosters stereotyping. See Fiske et al., *The Continuum Model*, supra note 36, at 242.

²⁷⁵ See Gerhard Daday & Beverly Burris, *Technocratic Teamwork: Mitigating Polarization and Cultural Marginalization in an Engineering Firm*, in *The Transformation of Work* 241, 254, 257 (Steven P. Vallas ed., 2001); see also Laurel Smith-Doerr, *Women's Work: Gender Equality vs. Hierarchy in the Life Sciences* 28–29, 147–48 (2004) (noting that women and minority scientists are more productive and do better in a network structures); Vicki Smith, *Employee Involvement, Involved Employees: Participative Work Arrangements in a White-Collar Service Occupation*, 43 *Soc. Probs.* 166, 177–78 (1996) (concluding that employee involvement programs benefit minorities by improving white-collar interactional skill sets).

2009]

Making Good on Good Intentions

1963

Job rotation schemes, which have been successful in many workplaces,²⁷⁶ work because they build a sense of competence, as well as autonomy and connectedness, making it more likely that people will internalize workplace nondiscrimination goals. These measures tend to enhance contact across groups, providing opportunities to form individual relationships that discourage group stereotyping. They also make more visible how the many parts of an operation contribute to a productive whole, which fosters a sense of interdependence and competence.

Accountability is widely praised as a way to reduce discrimination.²⁷⁷ A focus on motivation helps to fashion an accountability system that motivates people in a positive, constructive way. Accountability works when it motivates people to become more self-critical and to make more accurate, individuating (that is, non-stereotyping) decisions.²⁷⁸ When people know that their judgments of another person will be checked against the assessments of others whom they respect, they will want to form more careful judgments,²⁷⁹ and they will alter their attitudes accordingly.²⁸⁰ Similarly, people are more likely to shift attitudes in a positive direction when it is clear what is expected, as long as it does not require compromising basic convictions and preserves their sense of autonomy.²⁸¹ People need to know they will be accountable before

²⁷⁶ See, e.g., Ian M. Taplin, *Flexible Production, Rigid Jobs: Lessons from the Clothing Industry*, 22 *Work & Occupations* 412, 416–17 (1995).

²⁷⁷ Bielby, *supra* note 180, at 125–26.

²⁷⁸ Tetlock & Lerner, *supra* note 196, at 581–82.

²⁷⁹ See Pennington & Schlenker, *supra* note 196, at 1078.

²⁸⁰ Tetlock & Lerner, *supra* note 196, at 572 (“Accountability can affect not only what people say they think, but also how they actually do think.”); see also Fiske & Neuberg, *supra* note 195, at 41–42 (collecting studies that link the desire to form an accurate impression with an awareness that the impression will be shared); Pennington & Schlenker, *supra* note 196, at 1076–80 (describing the effects of knowing to which audience the study participants would have to defend their decision in a face-to-face meeting even when the anticipated meeting was cancelled before the participants wrote their decisions).

²⁸¹ Tetlock and Lerner note, for example, that the decisionmaker will not want to create dissonance with his own self-concept or to appear to others to be hypocritical or sycophantic. Tetlock & Lerner, *supra* note 196, at 575. It is also important that the people to whom they are accountable act consistently themselves with the norms they articulate. See Tony Simons et al., *Racial Differences in Sensitivity to Behavioral Integrity: Attitudinal Consequences, In-Group Effects, and “Trickle Down” Among Black and Non-Black Employees*, 92 *J. Applied Psychol.* 650, 658 (2007); Tetlock & Lerner, *supra* note 196 at 577.

they collect the evidence on which they must base their judgments, not after it is too late to guide their performance, when accountability will be experienced as negative and unfair.²⁸²

Accountability systems do not reduce discrimination when they motivate people to avoid criticism or to “become mired in self-justification.”²⁸³ Accountability systems that amount to second-guessing decisions after the fact do not work well²⁸⁴ because they undermine people’s sense of competence and autonomy and make them act defensively. When accountability systems appear to bribe decisionmakers or convey the message that without accountability they cannot be trusted to do the job, they undermine intrinsic motivation and thus are also less likely to de-bias decisionmaking.²⁸⁵ Giving people sufficient autonomy to experiment and take risks is a better way to build the trust necessary for them to invest themselves fully in the underlying goals.²⁸⁶

Positive feedback is generally viewed as autonomy- and competence-enhancing, and has been shown to enhance intrinsic motivation under some conditions; negative feedback, in contrast, tends to undermine people’s sense of effectiveness and thus intrinsic motivation.²⁸⁷ Accordingly, positive feedback has also been tied to a reduction in race stereotyping, and negative feedback to an increase in race stereotyping.²⁸⁸ For people to be internally motivated by positive feedback, however, they must feel responsible and engaged. Being praised for an outcome in which one played little role, or which was poorly executed, can create a sense of disso-

²⁸² Philip E. Tetlock, *Accountability and the Perseverance of First Impressions*, 46 *Soc. Psychol. Q.* 285, 290 (1983); Philip E. Tetlock & Jae Il Kim, *Accountability and Judgment Processes in a Personality Prediction Task*, 52 *J. Personality & Soc. Psychol.* 700, 706 (1987); Tetlock & Lerner, *supra* note 196, at 577–78.

²⁸³ Tetlock & Lerner, *supra* note 196, at 572.

²⁸⁴ *Id.* at 576–77.

²⁸⁵ See *id.* at 581–82. Thus, rewards that are unexpected are better than bribes because they make the individual feel more responsible for the achievement. See Deci et al., *Extrinsic Reward and Intrinsic Motivation in Education*, *supra* note 166, at 13–14.

²⁸⁶ Cf. Sturm, *The Architecture of Inclusion*, *supra* note 13, at 314 (describing how National Science Foundation builds relationship with grantees around trust, collaboration, risk-taking, and experimentation).

²⁸⁷ Deci & Ryan, *The “What” and “Why” of Goal Pursuits*, *supra* note 20, at 234–35 (citing studies); see also sources in *supra* note 202.

²⁸⁸ See *supra* text accompanying notes 68–72.

2009]

Making Good on Good Intentions

1965

nance, guilt, or inadequacy, rather than autonomy and competence. To have a positive effect on self-esteem, feedback should be genuinely affirming rather than controlling or manipulative.²⁸⁹ The form in which feedback is delivered matters. Informational feedback that affirms the need for accurate decisions reinforces people's sense of competence and responsibility. Didactic feedback is more likely to be experienced as judgmental or conditional, and thus as coercive.²⁹⁰ Research also shows that feedback that reinforces the purpose and value of the performance promotes the process of internalization better than feedback based solely on a do-as-you-are-told message.²⁹¹ Threats, deadlines, pressured evaluations, and imposed goals—where the justifications are not well understood—will tend to diminish intrinsic motivation, while opportunities for self-direction, and the acknowledgment of feelings tend to enhance it.²⁹²

Focusing on motivation helps to unpack various conundrums that have plagued scholars attempting to ascertain the best approaches to reducing unconscious bias. For example, while some research shows that guilt causes people to be less biased, other research shows the opposite. The discrepancy makes sense when guilt is connected to the quality of a person's motivation to be unbiased. People who have a well-internalized motivation to avoid prejudice respond to cues that they may be stereotyping with heightened attention, introspection, and effort to control stereotyping in future situations. In one study involving the review of law school applications, subjects who had high levels of internal motivation took more time to evaluate others after being told that their previous answers might have been influenced by the applicant's sexual orientation.²⁹³ Test subjects without such internalized stan-

²⁸⁹ See Deci & Ryan, The "What" and "Why" of Goal Pursuits, *supra* note 20, at 235.

²⁹⁰ See Deci et al., Extrinsic Reward and Intrinsic Motivation in Education, *supra* note 166, at 3, 12.

²⁹¹ Deci & Ryan, The "What" and "Why" of Goal Pursuits, *supra* note 20, at 238 (citing studies). So does acknowledging people's feelings that the activity is not interesting. *Id.*

²⁹² See Ryan & Deci, Self-Determination Theory and the Facilitation of Intrinsic Motivation, *supra* note 143, at 70 (citing studies).

²⁹³ Margo J. Monteith, Self-Regulation of Prejudiced Responses: Implications for Progress in Prejudice-Reduction Efforts, 65 *J. Personality & Soc. Psychol.* 469, 473, 477 (1993). They also responded more slowly to jokes about gays. *Id.* at 482.

dards made quicker evaluations in the next round, suggesting that the effect of the feedback was to make them more eager to avoid thinking about their prejudice, rather than to eliminate it.²⁹⁴ This research suggests that guilt can motivate people who have strongly internalized personal standards,²⁹⁵ but the threat associated with anticipated punishment or disapproval of others may fuel anger among those without such motivation.²⁹⁶ Anger and shame, in turn, increase the likelihood of stereotyping.²⁹⁷

The quality of a person's motivation can also help to explain otherwise puzzling evidence about the effects of confronting people about their bias. Researchers have found that the confrontation of those who utter stereotypes often generates apology, self-criticism, and subsequent positive changes in behavior among those who are motivated to avoid prejudice. But it makes other people defensive and heightens their stereotypic responses.²⁹⁸ One significant variable is the quality of the confrontation. Hostile confrontations are threatening, even to low-prejudiced people, because they impugn a person's sense of integrity.²⁹⁹ In reaction, people often conclude that "complainers" are hypersensitive.³⁰⁰ Low-key confrontations are less threatening and thus tend to be more productive.³⁰¹ The motivation of the person being confronted also matters.

²⁹⁴ *Id.* They also showed no moderation in their reaction to the homophobic jokes. *Id.* For other research developing the self-regulatory model of prejudice reduction, see Monteith et al., *supra* note 184; Margo J. Monteith & Aimee Y. Mark, Changing One's Prejudiced Ways: Awareness, Affect, and Self-Regulation, 16 *Eur. Rev. Soc. Psychol.* 113 (2005).

²⁹⁵ Devine et al., *supra* note 272, at 188–92.

²⁹⁶ *Id.* at 193–200.

²⁹⁷ See Fiske, Stereotyping, Prejudice, and Discrimination, *supra* note 44, at 390–91.

²⁹⁸ See Roy F. Baumeister et al., Personal Narratives About Guilt: Role in Action Control and Interpersonal Relationships, 17 *Basic & Applied Soc. Psychol.* 173, 187–88 (1995); Cheryl R. Kaiser & Carol T. Miller, Stop Complaining! The Social Costs of Making Attributions to Discrimination, 27 *Personality & Soc. Psychol. Bull.* 254 (2001); Monteith & Mark, *supra* note 294, at 146–47.

²⁹⁹ See Alexander M. Czopp et al., Standing Up for a Change: Reducing Bias Through Interpersonal Confrontation, 90 *J. Personality & Soc. Psychol.* 784, 785, 791 (2006) (reporting studies and further research indicating that even though hostile confrontations produced negative evaluations of the confronter, they also seemed to curb subsequent stereotyping); Monteith & Mark, *supra* note 294, at 147.

³⁰⁰ See Kaiser & Miller, *supra* note 298, at 261.

³⁰¹ See Czopp & Monteith, *supra* note 1, at 541. According to one study, confrontations of whites by other whites are less expected, less threatening, and more trustworthy, and thus more likely to promote a processing of the message. *Id.* at 534, 542; see

2009]

Making Good on Good Intentions

1967

Those who have internalized a desire to avoid discrimination are able to respond to confrontation positively and with a willingness to correct their behavior.³⁰² For people with high levels of prejudice, however, even relatively gentle encounters can be threatening and unproductive. These people are more likely to perceive confronters as unreasonable and thus to respond with hostility and self-justification.³⁰³

The conflicting data about the effects of affirmative plans also make more sense when motivational principles are considered. Numerous studies suggest that affirmative action programs devalue or stigmatize real and apparent beneficiaries.³⁰⁴ Yet the impact is less, and the acceptance by other people can be greatly enhanced, if programs are explained well,³⁰⁵ especially if they can be justified in terms of social norms and business rationales.³⁰⁶ Policies that focus on recruitment of underrepresented groups to hire good people

also Richard E. Petty et al., Individual Versus Group Interest Violation: Surprise as a Determinant of Argument Scrutiny and Persuasion, 19 Soc. Cognition 418, 430 (2001) (noting that people are considered more trustworthy when they advocate positions against their own self-interest, but not when they advocate positions against group interest). Still, some researchers conclude that confrontations may have positive long-term effects “through a sustained, strategic, and collective effort,” even if they cause short-term resistance. Czopp & Monteith, *supra* note 1, at 541.

³⁰² Czopp & Monteith, *supra* note 1, at 540–42.

³⁰³ *Id.* at 541. “High-prejudice people may be more influenced by [appeals to] universal norms of fairness and egalitarianism.” *Id.* at 542. Both high- and low-prejudice people felt more guilty and more uncomfortable when confronted about a biased response against blacks than when the response was against women. *Id.* at 541–42.

³⁰⁴ See, e.g., Madeline E. Heilman & Brian Welle, Disadvantaged by Diversity? The Effects of Diversity Goals on Competence Perceptions, 36 J. Applied Soc. Psychol. 1291, 1292 (2006); Kidder et al., *supra* note 271; Miriam G. Resendez, The Stigmatizing Effects of Affirmative Action: An Examination of Moderating Variables, 32 J. Applied Soc. Psychol. 185, 202 (2002).

³⁰⁵ See Ayres & Braithwaite, *supra* note 160, at 50 (“Just as strong external incentives retard internalization, using reasoning in preference to power-assertion tends to promote it.”); David A. Harrison et al., Understanding Attitudes Toward Affirmative Action Programs in Employment: Summary and Meta-Analysis of 35 Years of Research, 91 J. Applied Psychol. 1013, 1030 (2006) (conducting a meta-analysis of studies highlighting the importance of people understanding reasons for affirmative action plans); Orlando C. Richard & Susan L. Kirby, Women Recruits’ Perceptions of Workforce Diversity Program Selection Decisions: A Procedural Justice Examination, 28 J. Applied Soc. Psychol. 183, 184–87 (1998) (reporting that women given justification for diversity hiring suffer less negative self-perceptions than when given no justification); see also *supra* text accompanying note 159.

³⁰⁶ See Kidder et al., *supra* note 271, at 93.

1968

Virginia Law Review

[Vol. 95:1893]

and retain employees who reflect the diversity of their customer base are better accepted among both beneficiaries and potential opponents of policies.³⁰⁷ In contrast, measures that appear to favor members of protected groups at the expense of identifiable, seemingly deserving majority workers and are justified simply by the underrepresentation of the target group tend to provoke resistance to affirmative action more generally.³⁰⁸ This resistance was apparent in the highly negative public response to the City of New Haven's decision not to certify the results of a firefighters' promotion exam under which no black and only two Hispanic firefighters could have obtained promotions.³⁰⁹ From the perspective of motivation analysis, although New Haven never should have used a standardized, multiple-choice test to assess leadership potential,³¹⁰ once the city proceeded to use the test, it should have accepted the results and then simply discontinued future use of the test. The misunderstandings about affirmative action generated by the *Ricci v. DeStefano*³¹¹ decision—misunderstandings legitimized by the Supreme Court's decision in the case—have probably set back the

³⁰⁷ See Russell Cropanzano et al., Organizational Justice and Black Applicants' Reactions to Affirmative Action, 90 *J. Applied Psychol.* 1168, 1179–81 (2005); Heather Golden et al., Reactions to Affirmative Action: Substance and Semantics, 31 *J. Applied Soc. Psychol.* 73, 77, 80–82 (2001) (concluding that plans perceived as monitoring plans are better accepted than preferential hiring); Harrison et al., *supra* note 305, at 1020–28 (highly prescriptive affirmative action plans create greater resistance); Kidder, *supra* note 271, at 85, 88–89 (business priority justification associated with greater support toward affirmative action than concern about meeting affirmative action goals to adhere to EEOC guidelines for minority representation).

³⁰⁸ See Harrison et al., *supra* note 305, at 1020–28 (concluding that the rationale of better representation of underrepresented groups decreased support for affirmative action, especially among those who perceive that the plan is not in their self-interest, and when the plan is highly prescriptive).

³⁰⁹ See, e.g., Posting of Barry Friedman to The Plank, <http://www.tnr.com/blog/the-plank/why-should-ricci-should-frighten-democrats> (June 30, 2009, 14:19 EST) (arguing that *Ricci*, because of its sympathetic plaintiff-victim, revives Americans' ambivalence toward affirmative action and thus may weaken their commitment to it); see also Posting of Ed Kilgore to FiveThirtyEight, <http://www.fivethirtyeight.com/2009/06/obama-sptomayor-and-affirmative-action.html> (June 30, 2009, 17:09 EST) (showing that 65 percent of public support white firefighters in suit against New Haven, even though a Quinnipiac poll shows that 63 percent support affirmative action as long as rigid quotas are ruled out) (URL correct as displayed).

³¹⁰ See Lani Guinier & Susan Sturm, Trial by Firefighters, *N.Y. Times*, July 11, 2009, at A17.

³¹¹ 129 S. Ct. 2658 (2009).

2009]

Making Good on Good Intentions

1969

support for affirmative action that is critical to the reduction of implicit workplace bias.

More generally, motivation principles help to illuminate current debates about the significance of structural reform in the workplace. Some argue that changing institutional structures is more effective than changing individuals or the law.³¹² Others argue that internal reform tends to be co-opted by risk-management principles, amounting to “symbolic responses” without advancing the goals of equality and inclusion.³¹³ A major contribution of Susan Sturm’s research has been a framework for distinguishing the reforms that have the potential to transform institutions from those likely to simply perpetuate existing patterns. One promising strategy, she argues, is the development of collaborative problem-solving mechanisms through which non-judicial actors can promote diversity in the workplace.³¹⁴ Research by Sturm, Frank Dobbin, and Alexandra Kalev has also greatly improved understanding of the

³¹² See, e.g., Green & Kalev, *supra* note 34; Sturm, *Second Generation Employment Discrimination*, *supra* note 13, at 461–63; Sturm, *The Architecture of Inclusion*, *supra* note 13, at 249.

³¹³ See, e.g., Susan Bisom-Rapp, *Bulletproofing the Workplace: Symbol and Substance in Employment Discrimination Law Practice*, 26 Fla. St. U. L. Rev. 959, 962 (1999) (noting that preventative institutional practices undertaken to limit legal liability “may mask rather than eliminate some discriminatory decisions”); Lauren B. Edelman, *Law at Work: The Endogenous Construction of Civil Rights*, *in Handbook of Employment Discrimination Research: Rights and Realities* 337, 340–45 (Laura Beth Nielsen & Robert L. Nelson eds., 2005) (arguing that law is transformed by the organizational institutions that it is designed to control); Lauren B. Edelman, *Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law*, 97 Am. J. Soc. 1531, 1542–43 (1992) (“symbolic responses” occur when employers influence the meaning of employment discrimination laws through the internal procedures and structures they develop to comply with those laws); Lauren B. Edelman & Mark C. Suchman, *When the “Haves” Hold Court: Speculations on the Organizational Internalization of Law*, 33 L. & Soc’y Rev. 941, 985 (1999) (repeat player organizations annex law for their own purposes); Kimberly D. Krawiec, *Cosmetic Compliance and the Failure of Negotiated Governance*, 81 Wash. U. L. Q. 487, 487 (2003) (internal compliance mechanisms “may largely serve a window-dressing function that provides both market legitimacy and reduced legal liability”); see also Bisom-Rapp, *An Ounce of Prevention*, *supra* note 267 at 27–29 (courts increasingly treat education and training programs as evidence of nondiscrimination, which represents acceptance of form over substance).

³¹⁴ See Susan Sturm, *Law’s Role in Addressing Complex Discrimination*, *in Handbook of Employment Discrimination Research: Rights and Realities* 35 (Laura Beth Nielsen & Robert L. Nelson eds., 2005); Susan Sturm, *Second Generation Employment Discrimination*, *supra* note 13.

1970

Virginia Law Review

[Vol. 95:1893]

kinds of structural reforms that are most likely to make a difference in reducing bias. This research concludes that responsibility for diversity should be spread across the institution rather than focused in a single individual or administrative office; that in-house experts have advantages over outside consultants in building institutional diversity goals; that top management should be both diverse and committed to diversity; and that use of positive program incentives is more productive than legal compliance mechanisms, such as EEOC charges and lawsuits.³¹⁵

Principles of internal motivation help to explain these research findings and suggest ways to apply them more effectively. Each of the successful strategies support employee autonomy, competence, and relatedness. They rely on clear norms and provide opportunities for people to develop. Leadership and ownership are both critical. For institutional goals to have salience and credibility, the institution must reflect those values from the top. Employees can tell the difference between company leaders who model commitment to a diverse and inclusive workplace and those who act primarily from an aversion to legal liability.³¹⁶ Those who model commitment motivate others to do the same; those acting from an aversion to legal liability motivate cynicism and, at best, rote compliance. In turn, opportunities for people to “freely process and endorse transmitted values and regulations”³¹⁷ and to experience personal growth, feeling valued, and engagement give them a sense of ownership or what Susan Sturm calls institutional citizenship.³¹⁸

³¹⁵ See Frank Dobbin & Alexandra Kalev, *The Architecture of Inclusion: Evidence From Corporate Diversity Programs*, 30 *Harv. J.L. & Gender* 279, 292, 294–95, 300–01 (2007) (corporate context); Susan Sturm, *The Architecture of Inclusion*, *supra* note 13, at 247, 251, 300, 312–13 (in educational setting).

³¹⁶ In their study of the characteristics that make diversity programs effective in the corporate world, Professors Frank Dobbin and Alexandra Kalev identify strong leadership as a key component. See Dobbin & Kalev, *supra* note 316, at 295.

³¹⁷ Deci & Ryan, *The “What” and “Why” of Goal Pursuits*, *supra* note 20, at 238.

³¹⁸ See Sturm, *The Architecture of Inclusion*, *supra* note 13, at 323–27. Sturm describes and analyzes the National Science Foundation’s ADVANCE program for increasing the number of women in science and engineering. The program uses institutional transformation grants to support institutional change through data collection, data analysis, self-study, integrated strategies for improvement, monitoring and program assessment, and sharing of best practices. See *id.* at 277–334; see also David A. Thomas & Robin J. Ely, *Making Differences Matter: A New Paradigm for Managing Diversity*, *Harv. Bus. Rev.*, Sept.–Oct. 1996, at 79, 85–87 (urging paradigm for inclu-

2009]

Making Good on Good Intentions

1971

Without such opportunities, a job is just a job, not a shared enterprise whose institutional values are worth internalizing. Likewise, measures that make responsibility for various institutional goals diffuse, rather than someone else's bailiwick, foster institutional buy-in and stronger potential attachment to those goals. Without such commitment, people come to attribute their own compliance to the coercion, rather than their own values—an attribution that, as this Article has explained, tends to become a self-fulfilling prophecy.

CONCLUSION

There are three basic approaches for trying to reduce race and gender discrimination in the workplace. We can (1) prohibit it; (2) design the workplace so that people cannot engage in it; and (3) change the preferences and intentions that lead to it. Legal scholarship, understandably, has focused primarily on strengthening the legal rules that prohibit discrimination and, more recently, on proposing structural changes to the workplace that may block it. Changes in attitudes and preferences have been considered the incidental benefits of legal prohibitions and structural reforms, not their main, or even appropriate, purpose. Yet workplace discrimination cannot be eliminated solely through the threat of legal liability or the reform of workplace institutions that this threat encourages. In fact, too much pressure, or the wrong kind, may actually make things worse.

Faced with the frustrations of knowing that implicit discrimination exists and that the law is not effective in stopping it, there is an impulse to want to make the law do more. Because the law can deter some discriminatory behavior, it is tempting to assume that *more law* can *deter more* of it; that rules that produce more positive outcomes for plaintiffs will deter behavior that now falls between the cracks; and that stricter laws will incentivize employers to try even harder to eliminate bias from the workplace. As to the risk of false positives, the ubiquity of implicit bias makes it also easy to as-

sive workplace that provides opportunities for personal development, makes workers feel valued, and creates an expectation of high performance).

1972

Virginia Law Review

[Vol. 95:1893]

sume that “the opportunities for underestimating the incidence of discriminatory behavior are much greater than the reverse.”³¹⁹

Against these assumptions, however, must be considered the importance of good intentions in bringing about change that law, alone, cannot compel. Further, responses to the phenomenon of implicit bias must be informed not only by what more rules we can imagine, but also by what we know now, and can learn in the future, about the effects of various rules, institutions, and strategies on the good intentions that motivate people to act in nondiscriminatory ways, even when the law lacks the power to compel them to do so.

³¹⁹ McGinley, *Viva La Evolucion!*, supra note 17 at 482.