

Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment

Linda Hamilton Krieger†
Susan T. Fiske††

“The first call of a theory of law is that it should fit the facts.”
- Oliver Wendell Holmes¹

INTRODUCING BEHAVIORAL REALISM

Although they serve different social functions and employ different methods and tools, both law and the empirical social sciences need, use, and produce theories of human behavior. But their respective relationships to these theories differ in significant ways, and for this reason, law and social science often stand in tension with each other when they meet in the courtroom or the case reporter.

For its part, law needs, uses, and produces theories of human behavior when judges elaborate constitutional or common law doctrines or interpret ambiguous statutory provisions that implicate human motivation, subjective experience, or choice. Legal actors (judges, jurors, administrative fact finders, dispute handlers, and disputants) also use behavioral theories when they evaluate, litigate, or adjudicate specific disputes, as they attempt, for example, to attribute causation, assess witness credibility, or determine

Copyright © 2006 California Law Review, Inc. California Law Review, Inc. (CLR) is a California nonprofit corporation. CLR and the authors are solely responsible for the content of their publications.

† Professor of Law, University of California at Berkeley, School of Law (Boalt Hall). This work was supported by the Radcliffe Institute for Advanced Study at Harvard University and the Russell Sage Foundation. Helpful research was provided by I. Alejandra Góchez. Thanks to Mahzarin Banaji, Anthony Greenwald, Jerry Kang, Richard Banks, Gary Blasi, Vicki Shultz, Christine Jolls, Cass Sunstein, Ian Ayres, Ronald Allen, Brian Leiter, Laura Beth Neilson, Robert Nelson, Philip Frickey, Daniel Farber, Jan Vetter, Paul Schwartz, and Gillian Lester, who provided helpful comments on earlier drafts, and to participants at conferences, workshops, and lectures at the American Bar Foundation, Stanford Law School, New York University School of Law, The University of Georgia School of Law, the Harvard University Department of Psychology, the Radcliffe Institute for Advanced Study, U.C. Berkeley's Boalt Hall and U. C. Berkeley's Goldman School of Public Policy.

†† Professor, Department of Psychology, Princeton University (Green Hall).

1. OLIVER WENDELL HOLMES, *THE COMMON LAW* 167 (1963).

whether particular evidentiary facts constitute persuasive circumstantial evidence of some ultimate fact at issue in a particular case.

Acceptance of judicial authority depends heavily on jurisprudential stability, continuity, and adherence to precedent. As a result, judges are understandably hesitant to endorse proposed changes to the unstated psychological models underpinning legal doctrine, particularly if this would require modifying the doctrine itself. Consequently, as this Article discusses, behavioral theory change² comes slowly to law, when it comes at all.

In contrast to law, the empirical social sciences construct and use theories of human behavior with a more accepting eye toward their possible emendation, supplementation, or even outright replacement. While change in the behavioral theories embedded in law is viewed, most sympathetically, as a periodic necessary evil fraught with institutional peril, theoretical revision in the empirical social sciences codes as progress. To be sure, scientific communities resist fundamental theory change, as historians of science have vividly described.³ Nonetheless, the point remains: empirical social scientists expect that as empirical investigation continues and progress is made, the behavioral theories structuring their fields will evolve. In short, empirical psychology progressively refines its descriptive theories of human behavior because it is the goal of empirical psychology to do just that.

It is not the goal of law, however, to refine behavioral theories—even the behavioral theories embedded within legal doctrines. It is the goal of law to structure public and private ordering, to provide mechanisms for efficacious dispute resolution, and, in the process, to safeguard popular perceptions of judicial legitimacy. While the enterprise we call empirical social science is fundamentally descriptive, law is fundamentally normative. So, as we will describe, when law uses behavioral theories, it does so not in the interests of the theories themselves, but in pursuit of other goals.

For these and other reasons, the behavioral theories embedded in legal doctrines often go unstated. Even when stated, they are often unexamined, and they are almost never empirically tested, except perhaps by a small cadre of empirical “law and” scholars whose articles judges seldom read.⁴

2. To clarify, when in this Article we use the term “theory change” in reference to law, we are not referring to changes in legal theories themselves (e.g. disparate impact or disparate treatment theory in discrimination cases, or the theory of constructive intent in tort). Rather, we are speaking of the behavioral theories used by judges in constructing, applying, or justifying those legal theories. Our focus is on the psychological metatheories underpinning legal theories, not the legal theories themselves.

3. See, e.g., THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* (3d ed. 1996) (1962).

4. A fair amount of circumstantial evidence supports the claim that judges rarely read legal scholarship. See, e.g., Louis J. Sirico, Jr., *The Citing of Law Reviews by the Supreme Court: 1971-1999*, 75 *IND. L.J.* 1009, 1009-10 (2000) (reporting a decline in Supreme Court citation of legal scholarship);

Sometimes, behavioral theories enter case law as mere rhetorical flourishes, used to justify legal decisions made for reasons having nothing to do with the empirical validity of the theories themselves. However, once embedded in published decisions, a behavioral theory can develop precedential legitimacy, and for that reason be difficult to modify, even if it is empirically unsound.

When litigants attempt to use social scientific theories in factual adjudication, rules of evidence provide at least some institutionalized gatekeeping role in scrutinizing those theories' validity.⁵ But there is no systematic method through which the validity of behavioral theories used in legal reasoning is tested. Behavioral theories can thus enter and remain embedded in legal doctrine long after they have been disconfirmed or superseded by advances in the empirical social sciences.

The resulting inconsistencies between the real world and the phenomenological models embedded in law can be highly problematic. As Professor Krieger has argued in earlier work on antidiscrimination law, discontinuities between jurisprudential models of intergroup bias and the real world phenomena those models purport to represent have serious negative effects. These negative effects include normative ambiguity, adjudicative inefficiency and inaccuracy, and perhaps even an exacerbation of the very intergroup tensions antidiscrimination laws were enacted to diffuse.⁶

As Oliver Wendell Holmes, Jr. wrote in 1881, "The first call of a theory of law is that it should fit the facts."⁷ In context, of course, Holmes was referring to descriptive theories of law; he was not constructing, at least at that point, a prescriptive theory of judicial decision making. More recently, however, as typified by Justice Souter's dissent in *United States v. Morrison*,⁸ Holmes's "first call" principle has emerged as a rallying cry against the premising of legal doctrines on inaccurate conceptions of real world conditions. Used in this way, Holmes's "first call" principle points to

Michael D. McClintock, *The Declining Use of Legal Scholarship by Courts: An Empirical Study*, 51 OKLA. L. REV. 659, 660 (1998) (finding a 47% decline in the use of legal scholarship by courts over the previous two decades, with the most notable decline in the 1990s); see also Harry T. Edwards, *The Growing Disjunction Between Legal Education and the Legal Profession*, 91 MICH. L. REV. 34 (1992) (complaining that much of interdisciplinary legal scholarship is impractical and useless to courts); Richard A. Posner, *Legal Scholarship Today*, 115 HARV. L. REV. 1314, 1324 (2002) (observing that "interdisciplinary legal scholarship is intended to be read by professors . . . rather than by practitioners (including judges)").

5. See discussion *infra* p. 1022 and note 81.

6. Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1238-41 (1995) [hereinafter Krieger, *The Content of Our Categories*].

7. HOLMES, *supra* note 1.

8. 529 U.S. 598, 655 (2000) (Souter, J., dissenting) (arguing that, in declaring the Violence Against Women Act unconstitutional, the majority's conception of the nature of twenty-first century interstate commerce and the nature of the state-federal relationship failed Holmes's basic test that a theory of law should "fit the facts").

a prescriptive principle of adjudication, related to but distinct from the legal realism Holmes presaged.

This new principle, which the contributors to this Symposium call "behavioral realism," holds that as judges develop and elaborate substantive legal theories, they should guard against basing their analyses on inaccurate conceptions of relevant, real world phenomena. In this respect, behavioral realism echoes naturalizing epistemology, which emerged in the late 1960s with the work of W.V. Quine⁹ and has since made preliminary forays into jurisprudence¹⁰ and evidence scholarship.¹¹

Naturalism's core claim, like that of behavioral realism, is methodological: philosophical inquiry should remain continuous with advances in the empirical sciences. From this proposition, it follows that the development of theories of human knowledge acquisition should not be a totally conceptual, *a priori* endeavor. Rather, epistemic theories, both descriptive and normative, should be periodically revisited and revised to incorporate evolving understandings, derived from the empirical sciences, of how human cognitive processes actually work.¹²

An epistemic theory, like a theory of adjudication, can be descriptive or normative, and both types of theories can be naturalizing.¹³ Descriptive epistemic theories are naturalizing when they produce positive models that are consistent with those validated by empirical research in the cognitive sciences. Normative epistemic theories are naturalizing when they are structured in such a way that, based on relevant scientific evidence regarding knowledge acquisition and belief formation, one can reasonably assume that people will actually have the ability to comply with the prescriptions the normative epistemic theory generates.¹⁴

The contention that descriptive theories of human knowledge acquisition should be revised in light of advances in the cognitive sciences seems almost self-justifying. If key elements of a descriptive epistemic theory have been empirically disconfirmed, the theory has lost any claim to descriptive accuracy it might otherwise have had.

With respect to prescriptive theories, however, naturalism presents a less obviously correct claim. What people actually tend to do does not necessarily equate to what people *should* do. In the end, however, even naturalism's normative claim is sound in the sense that one cannot expect much

9. W.V. QUINE, *Epistemology Naturalized*, in *ONTOLOGICAL RELATIVITY AND OTHER ESSAYS* 69 (1969).

10. For an overview of naturalizing jurisprudence, see Brian Leiter, *Naturalism and Naturalized Jurisprudence*, in *ANALYZING LAW: NEW ESSAYS IN LEGAL THEORY* 79 (Brian Bix ed., 1998).

11. See generally Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491 (2001) (arguing that developments in epistemology can provide a conceptual foundation for addressing common problems in the law of evidence).

12. ALVIN I. GOLDMAN, *EPISTEMOLOGY AND COGNITION* 53 (1986).

13. Alvin I. Goldman, *Epistemics: The Regulative Theory of Cognition*, 75 J. PHIL. 509 (1978).

14. See *id.* at 513.

regulatory work out of a prescriptive theory of human knowledge acquisition or belief formation that is based on faulty premises about how people actually think.¹⁵ As Allen and Leiter put the matter, “[W]e cannot craft epistemic norms (norms that would guide our acquisition of knowledge) without empirical information.”¹⁶

Behavioral realism in law stands for the proposition that legal theories, no less than their epistemological counterparts, both can and should be naturalizing. Behavioral realism, like naturalism, stands for the proposition that judges should not generate the behavioral theories sometimes used in the construction or justification of legal doctrine through a solely conceptual, *a priori* process. To the extent that legal doctrines rely on stated or unstated theories about the nature of real world phenomena, behavioral realism argues, those theories should remain consistent with advances in relevant fields of empirical inquiry. And where the real world phenomena relevant to a particular area of law concern human social perception, motivation, and judgment, the relevant domains of empirical inquiry with which legal theories should remain consistent include cognitive social psychology and the related social sciences.

In the context of antidiscrimination law, behavioral realism stands for the proposition that judicial models—of what discrimination is, what causes it to occur, how it can be prevented, and how its presence or absence can best be discerned in particular cases—should be periodically revisited and adjusted so as to remain continuous with progress in psychological science.

The same earlier-described arguments defending naturalizing epistemology support the claim that antidiscrimination law should be naturalizing, that is, behaviorally realistic. First, antidiscrimination law should be behaviorally realistic because a normative theory of nondiscrimination based on faulty premises about how and why decision makers treat people differently because of their social group status cannot realistically perform much normative work. Even if people *want* to conform their behavior to the norms underlying antidiscrimination law, full compliance with the law’s prescriptions is unlikely if the relevant legal doctrines fail to capture accurately how and why discrimination occurs, how targets respond to it, and what can be done to prevent it from occurring.¹⁷ Furthermore, antidiscrimination doctrine should be naturalizing, or behaviorally realistic, because a legal model of discrimination that fails to reflect accurately its real

15. *Id.*

16. Allen & Leiter, *supra* note 11, at 1494.

17. For example, in earlier work, Krieger demonstrated that the colorblindness approach to nondiscrimination will fail as a normative theory because it is based on an empirically inaccurate model of the psychological processes involved in biased intergroup perception and judgment. Linda Hamilton Krieger, *Civil Rights Perestroika: Intergroup Relations After Affirmative Action*, 86 CALIF. L. REV. 1251, 1276-93 (1998) [hereinafter Krieger, *Civil Rights Perestroika*].

world counterpart will prove inefficient and ineffective from a forensic standpoint. Ironically, one might criticize such legal theories as lacking field validity.

But law, of course, is not epistemology, and in interpreting statutes or constitutional provisions, or in developing or amplifying common-law doctrines, judges are constrained by prudential considerations that hold no claim on epistemologists. In construing statutes, for example, federal judges are prudentially constrained by the statute's text, the judge's role as an unelected official in a democratic society, and the principle of *stare decisis*.

Of course, in many situations, empirically testable claims play no role in the legal analysis grounding judicial interpretation of statutes or constitutional provisions, or in the elaboration of common-law rules. But in interpreting and applying ambiguous statutory provisions, in crafting legal rules to incentivize behavior consistent with a statute's purpose, or in adjudicating individual cases, judges necessarily draw on models of real world phenomena and incorporate those models into their legal reasoning. When these models represent empirically testable—or even tested—claims, behavioral realism maintains that judges should take reasonable steps, whether through the solicitation of expert testimony, amicus participation, or otherwise, to make sure they have the science right.¹⁸

Stated in this way, the core principle underlying behavioral realism seems obviously correct. But where psychological theories are involved, judicial compliance with this principle is harder than one might suppose. This is because judges, like most people, take for granted certain assumptions about how people behave and what motivates them. These assumptions seem self-evidently correct, even when they are wrong. For this reason, judges sometimes incorporate empirically testable social science claims into their legal reasoning without even noticing that they are doing so. As we discuss in Part I, judges, no less than litigants, are quick to recite and hesitant to question the behavioral theories that underpin established legal doctrines.

Slowly, however, this may be changing. As Thomas Ulen has described,¹⁹ the emergence of law and economics has played a major role in prodding legal scholarship—and, in certain instances, legal reasoning—

18. In response to those who would argue that judges cannot be expected to understand psychological science, or apply its insights when making choices among competing legal rules, one need only point to the enormous influence of economics on doctrinal questions in contract, securities regulation, antitrust, and even employment discrimination law. See RICHARD POSNER, *ECONOMIC ANALYSIS OF LAW* 19-22 (1992) (summarizing the influence of economics on law in a variety of subject-matter areas). It is difficult to justify a double standard that would assume that judges are able to distinguish good from bad economics but incapable of the same discernment in other social science disciplines.

19. Thomas S. Ulen, *The Unexpected Guest: Law and Economics, Law and Other Cognate Disciplines, and the Future of Legal Scholarship*, 79 CHI.-KENT L. REV. 403, 404-05 (2004).

toward a more scientific method of inquiry.²⁰ Over time, other social science disciplines, most notably sociology and cognitive psychology, have extended the empirical interdisciplinary turn that law and economics initiated. Within the academy, for example, the participation of social science disciplines beyond neoclassical economics gave rise to the empirical wing of the law and society school and to behavioral law and economics.

Despite their differences, these various “law and” schools share a common methodology: they all seek to identify empirically testable assumptions on which legal doctrine or policy initiatives are premised, analyze those assumptions using their respective disciplines’ methods and tools, and then use the results of that analysis to evaluate existing legal doctrines or policies and, where indicated, to suggest reforms.

As part of this broad movement, there has emerged in the past ten years a school of legal scholarship exploring the implications of insights emerging from psychological science for antidiscrimination law and policy.²¹ Taken as a whole, this scholarship chronicles the many ways in which

20. See discussion *infra* Part II.B (text accompanying notes 74-78).

21. A complete cataloging of the relevant literature would be impractical for an essay of this relative brevity, so we offer only a sampling of employment law related scholarship published during the past ten years: Krieger, *The Content of Our Categories*, *supra* note 6 (applying insights from early social cognition research to disparate treatment doctrine); Krieger, *Civil Rights Perestroika*, *supra* note 17 (using insights from social and cognitive psychology to evaluate various arguments for and against affirmative action); Amy L. Wax, *Discrimination as Accident*, 74 IND. L. J. 1129 (1999) (arguing against imposing Title VII liability for discrimination stemming from cognitive bias); Cynthia L. Estlund, *Working Together: The Workplace, Civil Society, and the Law*, 89 GEO. L.J. 1 (2000) (using insights from social psychology and, in particular, the contact effect, to evaluate social gains derived from workplace integration); Ann C. McGinley, !Viva La Evolución!: *Recognizing Unconscious Motive in Title VII*, 9 CORNELL J.L. & PUB. POL’Y 415 (2000) (extending Krieger’s earlier analysis of the implications of social cognition theory for proof of disparate treatment); Ian F. Haney-López, *Institutional Racism: Judicial Conduct and a New Theory of Racial Discrimination*, 109 YALE L.J. 1717 (2000) (drawing on new institutionalism and social cognition theory to explicate institutionalized forms of race discrimination in the selection of grand jurors); Martha Chamallas, *Deepening the Legal Understanding of Bias: On Devaluation and Biased Prototypes*, 74 S. CALIF. L. REV. 747, 748-53 (2001) (using cognitive social psychology to develop the concepts of devaluation and prototyping and applying them to the law of intentional discrimination); 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 (2001) (examining social science evidence regarding the effectiveness of diversity and anti-sexual-harassment training in preventing discrimination and harassment); Rebecca Hanner White & Linda Hamilton Krieger, *Whose Motive Matters?: Discrimination in Multi-Actor Employment Decision Making*, 61 LA. L. REV. 495 (2001) (applying insights from the social psychology of group decision making to disparate treatment cases involving multiple decision makers); Michelle A. Travis, *Perceived Disabilities, Social Cognition and “Innocent Mistakes,”* 55 VAND. L. REV. 481 (2002) (applying social cognition theory to the problem of disability discrimination); Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 HARV. C.R.-C.L. L. REV. 91 (2003) (arguing that insights from cognitive social psychology should be used to inform how disparate impact theory is interpreted and applied in employment-discrimination litigation); Tristin K. Green, *Targeting Workplace Context: Title VII as a Tool for Institutional Reform*, 72 FORDHAM L. REV. 659 (2003) (extending her earlier analysis and applying it to issues relating to remedies); Gary Blasi, *Advocacy Against the Stereotype: Lessons From Cognitive Social Psychology*, 49 UCLA L. REV. 1241

established civil rights jurisprudence is premised on models of social perception and judgment that have been significantly discredited by empirical work in social and cognitive psychology.

Reflecting this behavioral realist turn in civil rights scholarship, many scholars have drawn on advances in the empirical social sciences to demonstrate that what the law refers to as “intentional discrimination” can just as easily result from the uncontrolled application of implicit, unconscious, or automatic stereotypes and other subtle ingroup preferences as from the operation of conscious discriminatory designs. These scholars, who include both lawyers and social psychologists, endeavor to identify this broader set of mental processes that results in disparate treatment of members of negatively stereotyped or otherwise marginalized groups. In doing so, they generally advocate a causation-based, rather than an intent-based, understanding of the antidiscrimination principle.²² They also support an expansive application of disparate impact theory in cases involving subjective decision-making systems or other processes or criteria that tend to systematically deprive historically marginalized groups of employment opportunities.²³

Intellectual collaboration between lawyers and psychologists is nothing new in American civil rights law. Collaborations between civil rights lawyers, progressive legal scholars, and politically engaged social scientists have contributed to some of the most important advances in federal civil rights jurisprudence over the past half-century. The example that comes most readily to mind, of course, is the collaboration between the NAACP Legal Defense and Education Fund and the social scientists who authored and signed the *Social Science Statement* submitted to the Supreme Court in

(2002) (exploring the consequences of cognitive biases in litigation involving members of stereotyped groups); Lu-in Wang, *Race as Proxy: Situational Racism and Self-Fulfilling Stereotypes*, 53 DEPAUL L. REV. 1013 (2004) (using insights from social cognition theory to argue that individual disparate treatment adjudications are a poor policy tool for addressing modern forms of race discrimination); Melissa Hart, *Subjective Decisionmaking and Unconscious Discrimination*, 56 ALA. L. REV. 741 (2005) (applying insights from social cognition theory to the problem of subjective employment decision making systems).

22. White & Krieger, *supra* note 21; McGinley, *supra* note 21; Michael Selmi, *Response to Professor Wax, Discrimination as Accident: Old Whine, New Bottle*, 74 IND. L.J. 1233 (1999); Evan Tsen Lee & Ashutosh Bhagwat, *The McClesky Puzzle: Remedying Prosecutorial Discrimination Against Black Victims in Capital Sentencing*, 1998 SUP. CT. REV. 145 (1998); Mary Ellen Maatman, *Choosing Words and Creating Worlds: The Supreme Court's Rhetoric and Its Constitutive Effects on Employment Discrimination Law*, 60 U. PITT. L. REV. 1 (1998); Michael Selmi, *Proving Intentional Discrimination: The Reality of Supreme Court Rhetoric*, 86 GEO. L.J. 279 (1997); Michael J. Zimmer, *The Emerging Uniform Structure of Disparate Treatment Discrimination Litigation*, 30 GA. L. REV. 563, 600-09 (1996); Krieger, *The Content of Our Categories*, *supra* note 6, at 1242-43; David A. Strauss, *Discriminatory Intent and the Taming of Brown*, 56 U. CHI. L. REV. 935 (1989).

23. See, e.g., Tristin K. Green, *Work Culture and Discrimination*, 93 CALIF. L. REV. 623 (2005); Green, *Targeting Workplace Context*, *supra* note 21; Robert Belton, *Title VII at Forty: A Brief Look at the Birth, Death, and Resurrection of the Disparate Impact Theory of Discrimination*, 22 HOFSTRA LAB. & EMP. L.J. 431 (2005).

Brown v. Board of Education.²⁴ The American Psychological Association's amicus participation in *Price Waterhouse v. Hopkins*,²⁵ following on the trial testimony of Susan T. Fiske, played a significant role in the Court's endorsement of mixed motive theory in Title VII cases.

While these collaborations can reasonably be characterized as successful, they could not accurately be described as uncontroversial. On the contrary, they have long raised thorny and still hotly contested questions about the proper role of social science in law generally, and in discrimination law more specifically. *Brown's* footnote eleven unleashed harsh criticism, not only of Kenneth and Mamie Clark's doll studies, but of the very notion that expert social psychological testimony had any proper role in constitutional jurisprudence.²⁶ Since the Supreme Court's 1993 decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,²⁷ "social framework"²⁸ and related social science testimony has been met with renewed interest, and skepticism, from both commentators and courts.²⁹

Those who criticize the use of social science in shaping legal doctrine have long argued that normative legal principles and traditional tools of statutory and constitutional interpretation—not social science theories—should guide substantive lawmaking.³⁰ Judges, these critics demand,

24. See generally JOHN P. JACKSON, JR., SOCIAL SCIENTISTS FOR SOCIAL JUSTICE: MAKING THE CASE AGAINST SEGREGATION 109-82 (2001) (examining the role of social scientists in the *Brown* litigation).

25. 490 U.S. 228 (1989) (endorsing the use of mixed motive theory in Title VII disparate treatment cases).

26. The classic statement is Herbert Wechsler, *Toward Neutral Principles of Constitutional Law*, 73 HARV. L. REV. 1, 31-34 (1959) (criticizing the use of social science research in the Supreme Court's decision in *Brown v. Board of Education*). Other critical treatments include Monroe Berger, *Desegregation, Law, and Social Science*, 23 COMMENTARY 471, 476 (1957) (noting that legal protection from segregation should not depend on empirical findings of actual harm); Edmond Cahn, *Jurisprudence*, 30 N.Y.U. L. REV. 150, 157-58 (1955) (suggesting that empirical demonstrations are a "flimsy foundation" on which to rest equality rights).

27. 509 U.S. 579 (1993).

28. The use of the phrase "social framework" testimony in this context originates from Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559 (1987). Walker and Monahan describe social framework testimony as the use of social science to provide fact finders with a pattern of "general research results . . . used to construct a frame of reference or background context for deciding factual issues crucial to the resolution of a specific case." For a recent discussion of the use of social framework testimony, see, e.g., William T. Bielby, *Can I Get a Witness? Challenges of Using Expert Testimony on Cognitive Bias in Employment Discrimination Litigation*, 7 EMP. RTS. & EMP. POL'Y J. 377 (2003).

29. See Timothy Zick, *Constitutional Empiricism: Quasi-Neutral Principles and Constitutional Truths*, 82 N.C. L. REV. 115 (2003) (arguing that empirical methods are ill suited to the discovery of constitutional meaning); see also *Kotla v. Regents of the Univ. of Cal.*, 8 Cal. Rptr. 3d 898 (Cal. Ct. App. 2004) (holding that social framework testimony about stereotyping in employment-discrimination case was not useful to the trier of fact because the nature and effects of stereotypes are within people's ordinary common knowledge).

30. See PHILIP BOBBITT, *CONSTITUTIONAL FATE: THEORY OF THE CONSTITUTION* 6, 94 (1982).

“should abandon the practice of basing their decisions on the basis of empirical propositions.”³¹

Identifying the proper role of social science in substantive lawmaking is an enormously complex theoretical undertaking, and it is not our purpose to resolve here the many vexing questions that undertaking represents. Our aim is far more modest. Using Title VII jurisprudence as a reference point, we argue here that those who criticize the use of insights from empirical social psychology in shaping or applying antidiscrimination doctrines are overlooking one extremely important point—social psychology is already there. In discrimination cases, as elsewhere, judges are constantly using “intuitive” or “common sense” psychological theories in the construction and justification of legal doctrines and in their application to specific legal disputes. A psychologically trained eye can spot these intuitive psychological theories all across Title VII’s doctrinal landscape. Behavioral realism is not a jurisprudential innovation; it is a jurisprudential corrective. In discrimination law, there already is, and there has long been, an “intuitive psychologist behind the bench.”³²

The problem is, as Stanford social psychologist Lee Ross observed many years ago, the “intuitive psychologist” has significant shortcomings.³³ When subjected to empirical scrutiny, “common sense” theories of how people perceive and judge themselves and others in their social environment often turn out to be wrong. Behavioral realism, understood as a prescriptive theory of judicial decision making, addresses this problem by proposing that, before judges use lay or “common sense” psychological theories in their legal analysis, they should take reasonable steps to ensure that those theories are valid.

We do not mean to suggest that behavioral realism can or should replace other jurisprudential principles. As a prescriptive principle of adjudication, behavioral realism must cohere and coexist with, and sometimes yield to, other jurisprudential norms. So, for example, even if a statutory provision was based on erroneous empirical assumptions, if it was textually unambiguous, its meaning made clear by legislative history and applicable canons of statutory interpretation, behavioral realism would provide no warrant for changing it by judicial fiat. The statute’s erroneous empirical premises might provide an argument in favor of its amendment, but that argument would properly be made to the legislature, not the courts.

We also do not argue here that where a well-established precedent was based upon a faulty empirical claim, that fact, in and of itself, would

31. David M. O’Brien, *The Seduction of the Judiciary: Social Science and the Courts*, 64 JUDICATURE 8, 11 (1980).

32. Linda Hamilton Krieger, *The Intuitive Psychologist Behind the Bench: Models of Gender Bias in Social Psychology and Employment Discrimination Law*, 60 J. SOC. ISSUES 835 (2004).

33. Lee Ross, *The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process*, in COGNITIVE THEORIES IN SOCIAL PSYCHOLOGY 337 (Leonard Berkowitz ed., 1978).

justify the precedent's overruling. When considering whether to override the principle of *stare decisis*, a court must take numerous considerations into account, including, among others, whether the existing rule has proven practically unworkable, the nature and extent of the reliance interests it has generated, and the costs of an overruling to the Court's perceived legitimacy, among others.³⁴

Moreover, even the best insights from the empirical social sciences cannot supply the normative principles needed for substantive lawmaking or resolve the conflicts between competing norms and interests so often implicated in the legislative and judicial processes. Law, at root, is normative. Empiricism, at root, is descriptive, and it must remain so to fulfill its proper function.

Nonetheless, a place for behavioral realism remains. Statutory terms and prior judicial precedents sometimes *are* ambiguous. Precedents often underdetermine legal outcomes. The common law process leaves substantial room for the judicial elaboration of substantive law. Moreover, even when the contours of the applicable substantive law are clear, in adjudicating particular cases, judges are often called upon to draw inferences from evidence, or to determine whether the inferences a trier of fact might draw from evidence would be reasonable.

Even conceding that normative legal analysis, and not empiricism, must supply the fundamental principles with which judges work, once the law's normative goals have been specified, they must be operationalized through particular legal doctrines and policies. When those doctrines and policies are based on faulty models of relevant social phenomena, the law's ability to advance its normative agenda will be compromised. In short, when the doctrines and policies designed to operationalize normative legal principles are premised on empirically testable assumptions about human social thought or behavior, it is hard to argue with the proposition that insights from the empirical social sciences have an important role to play in ensuring that those assumptions are sound.

In this Article, we seek to illustrate and advance behavioral realism in law by applying its methods to the problem of defining and identifying discriminatory motivation in Title VII individual disparate treatment cases. We begin in Part I by examining the roles that psychological theories play in law and in empirical social psychology, and by describing the processes through which each discipline develops and evaluates—or fails to evaluate—the psychological theories it employs. Through this analysis, we hope to show that the most common objection to behavioral realism in law—namely, that legal analysis should be based on normative principles rather than social science theories—is misplaced. Behavioral realism does not inject social science theories into legal reasoning. It merely provides a clear

34. *Planned Parenthood v. Casey*, 505 U.S. 833, 854-55 (1992).

and constructive process for recognizing, evaluating, and, where necessary, modifying social science theories that are already there.

In Part II, we illustrate how behavioral realism can function as a critical tool in doctrinal analysis. Specifically, we explore how a behavioral realist approach can be used to inform the analysis of two hotly disputed issues in individual disparate treatment doctrine: the same actor inference and the honest belief rule. Through this analysis, we show specific evidence of the claim we made in Part I—that behavioral realism does not inject social science theories into legal doctrine. Rather, it simply identifies and subjects to scrutiny the unexamined social science theories, many of them plainly invalid, that are already there. Finally, in Part III, we offer a positive theory of discriminatory motivation, informed both by behavioral realist methods and by traditional—even conservative—approaches to statutory interpretation. Through this discussion, we seek to show that, far from representing a radical innovation, the behavioral realist reading of the disparate treatment principle we propose is fully consistent with Supreme Court precedent, textualist methods of statutory interpretation, and principles of judicial restraint.

I

THEORIES OF HUMAN BEHAVIOR IN LAW AND EMPIRICAL PSYCHOLOGY

In developing, justifying, and applying the substantive law in any arena implicating motivation, judgment, or choice, judges frequently employ theories of human behavior. Both stated and unstated behavioral theories work their way into the substantive law as judges define ambiguous terms in statutes or case precedents; as they identify and elaborate upon the essential elements of proof constituting common law causes of action; as they develop judicially sanctioned inferences or presumptions; and as they articulate moral or policy justifications for new or existing legal doctrines.

Empirical social psychology also uses theories of human behavior. In that discipline, theories are used to explain patterns of experimental observations and to structure future investigation. However, unlike appellate judges, who are relatively free to incorporate psychological theories into their legal analysis without having to establish their validity, psychological scientists constantly subject their field's theories to empirical scrutiny under consistent and widely agreed-upon methods and evaluative criteria. This is not the situation in jurisprudence, a field that has yet to develop well-defined methods and criteria for evaluating the behavioral theories used in substantive lawmaking. Even in a precedential decision, a judge who incorporates a particular lay theory of how stereotypes influence social judgment does not need to show, and probably would never even think to ask, whether that theory had predictive validity, field validity, or any validity at all. As we will attempt to demonstrate here, these differences in

how law and psychological science develop and evaluate behavioral theories have significant and sometimes negative consequences for law. Behavioral realism seeks to address these consequences.

A. Behavioral Theories in Law

Law uses theories of human behavior for a number of distinct purposes. Judges use such theories when they interpret the meaning of ambiguous statutory provisions that implicate human perception, judgment, or choice. Behavioral theories also figure prominently in the development of common law theories of liability, as judges identify and elaborate the essential elements of proof and defense that constitute the various common law crimes and civil causes of action. Behavioral theories work their way into constitutional law, as well, most notably as judges attempt to justify established or novel constitutional doctrines. Trial and administrative judges, litigants, jurors, and dispute handlers also use behavioral theories when they evaluate or adjudicate specific disputes, as they attempt, for example, to attribute causation, assess witness credibility, or determine whether particular evidentiary facts constitute persuasive circumstantial evidence of some ultimate fact at issue. And finally, judges, legislators, and administrative regulators use behavioral theories when they formulate legal rules in an attempt to structure behavior through individual or organizational incentives. In the discussion that follows, we elaborate on each of these uses of behavioral theories in the judicial process.

1. The Use of Behavioral Theories in Statutory Interpretation

When judges are called upon to interpret and apply statutes that bear on social perception, judgment, and choice, they need theories of human behavior in order to interpret ambiguous statutory terms and apply them to particular fact patterns. Judicial explication of terms that connote phenomena implicating cognition, affect, incentives, and choice necessarily rely upon psychological models of how people go about making sense of themselves and others in the social world. Because judges need psychological theories to interpret statutes, it should come as no great surprise that judges produce and use those theories, and that the theories either directly or indirectly influence how the law develops over time.

For example, when enacted in 1964, Title VII made discrimination based on race, color, sex, national origin, or religion unlawful.³⁵ But the statute nowhere defined what “discrimination” meant. Then, when Congress passed the Civil Rights Act of 1991, it amended Title VII to provide that actionable discrimination occurred whenever someone’s race, color, sex, national origin, or religion was a “motivating factor” in an

35. Section 703(a)(1), Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e-2(a)(1).

employment decision.³⁶ Here again, Congress did not explain what it meant by a “motivating factor.”

As a result, the courts turned to how they themselves had previously defined the term. In *Price Waterhouse v. Hopkins*,³⁷ a 1989 Supreme Court decision endorsing the use of “mixed motive” analysis in Title VII disparate treatment cases, the plurality relied on a “motivating factor” analysis. In describing for the plurality what it means for gender bias to “motivate” a challenged employment decision, Justice Brennan wrote:

In saying that gender played a motivating part in an employment decision, we mean that, if we asked the employer at the moment of the decision what its reasons were and if we received a truthful response, one of those reasons would be that the applicant or employee was a woman.³⁸

This description reflects two “common sense” theories about the nature of discriminatory motivation. In speaking of the decision maker providing a “truthful” (as opposed to an “accurate”) response, this description reflects an unstated assumption that, when disparate treatment discrimination occurs, the discriminator is consciously aware, “at the moment of decision,”³⁹ that he or she is discriminating. As Professor Krieger described in earlier work,⁴⁰ the implications of which will be extended later in this article, these two lay psychological theories—the belief in transparent mental processing, and the modeling of perception and decision making as two discrete processes—have not withstood empirical scrutiny. Decision makers are often *not* aware of the impact of a target’s social group membership on their judgments, and those biased judgments are often formed quite early in the social perception process, long before the moment that a decision about the target person is made.⁴¹

As this illustration suggests, it is not possible for judges to interpret what Congress meant when it used the term “discrimination” or the phrase “motivating factor” in Title VII without applying psychological theories about how human social perception, motivation, and judgment work—that is, psychological theories about when people are discriminating and when they are not.

2. *The Use of Psychological Theories in Civil Adjudication*

In civil adjudication, judges and other legal actors employ models of human behavior in the production and justification of abstract legal

36. Section 703(m), Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e-2(m).

37. 490 U.S. 228 (1989).

38. *Id.* at 250.

39. *Id.*

40. Krieger, *The Content of Our Categories*, *supra* note 6.

41. See discussion *infra* pp. 60-62 and notes 146-152.

doctrines, and in the application of those doctrines to concrete cases. In either situation, judicial reliance on psychological models may be expressed directly, or it may function as an unstated, taken-for-granted conception of the various social and psychological phenomena implicated by a particular area of law.

At least in the Anglo-American system, law seeks to regulate individual and organizational behavior through the elaboration and application of doctrinal structures variously referred to as “causes of action,” “remedial theories,” or “claims for relief.” If a person feels wronged, and desires legal redress for that wrong, she must fit the facts of her case into at least one legally cognizable claim for relief.

A claim for relief can usefully be understood as a type of schema, which we might call a “claim schema.” Every claim schema comprises a particular set of narrative and analytical elements. On the narrative side, each claim schema is associated with a script—a prototypic narrative in which prototypic characters, acting from prototypic motivations, behave in prototypic ways with prototypic consequences for a prototypic victim, who becomes a prototypic claimant. From an advocacy perspective, the narrative resonance of one’s case is key to its success. As Anthony Amsterdam and Jerome Bruner observe, “Law lives on narrative, for reasons both banal and deep.”⁴²

But narrative flow is not all that matters; law has a formal analytical side as well. In this regard, any claim schema can be parsed into a set of “essential elements,” legal analytic components that must be discernable in the narrative to form a legally cognizable claim. To establish entitlement to legal relief, a claimant must introduce admissible evidence sufficient to establish the existence of each essential element of one or more legally recognized causes of action.

For example, to establish a right to recover for fraud, a plaintiff must establish five essential elements constituting the fraud cause of action: 1) the making of a false or misleading representation to the victim; 2) with actual or constructive knowledge of its falsity or misleading nature; 3) intent to defraud; and 4) actual reliance on the victim’s part, which reliance; 5) redounds to the victim’s detriment.⁴³ To establish a claim for individual

42. ANTHONY G. AMSTERDAM & JEROME BRUNER, *MINDING THE LAW: HOW COURTS RELY ON STORYTELLING, AND HOW THEIR STORIES CHANGE THE WAYS WE UNDERSTAND LAW - AND OURSELVES* 110 (2000).

43. The precise articulation of these essential elements differs slightly from jurisdiction to jurisdiction, but they correspond in general substance. So, for example, to establish concealment, fraud, or misrepresentation, under California law a plaintiff must prove that: 1) the defendant made a representation as to a past or existing material fact; 2) the representation was false; 3) the defendant knew that the representation was false when made; 4) the defendant made the representation with an intent to defraud the plaintiff—with the purpose of inducing the plaintiff to rely on it; 5) the plaintiff was unaware of the falsity of the representation and acted on the assumed truth of the representation;

disparate treatment discrimination under Title VII, a plaintiff must establish through the introduction of admissible evidence: 1) that he or she is a member of a protected group; 2) that he or she was subjected to some negative employment decision; and 3) that his or her protected-group status was a motivating factor in that decision.⁴⁴ In either of these examples, and indeed in any legal case, a failure of proof on any one of the claim's essential elements results in a judgment for the defendant.⁴⁵

Lay psychological theories necessarily find their way into law at the level of doctrinal construction as courts define and describe the essential elements of proof and defense constituting legally cognizable crimes or civil claims for relief. For example, theories of intentionality figure centrally in the substantive criminal law, where crimes are defined largely by the subjective state of mind with which a particular act was performed. In this regard, for example, recklessness is distinguished from intent, and mere intent from intent with malice aforethought. Criminal adjudication requires the trier of fact to discern, usually through the drawing of inferences from circumstantial evidence, the subjective state of mind with which the crime's *actus reus* was performed. For this reason, criminal legal doctrine necessarily reflects and reifies a particular theory of the mind. So, too, with respect to the law of torts, where the same act and resulting harm is accorded profoundly different legal significance according to the state of mind with which an ultimately harmful act is performed, from mere thoughtlessness to malicious injurious purpose.

The incorporation of a particular essential element into a crime or civil cause of action may in and of itself reflect an implicit psychological theory. For example, we assume (and we do not mean to suggest that this assumption is incorrect), that intent to kill is practically as well as theoretically distinguishable from malice aforethought. The substantive law of fraud implicitly assumes that "knowledge of falsity" is psychologically distinct and empirically distinguishable from the spontaneous, subjective construal of ambiguous information in a self-serving manner, a well-documented cognitive process that often occurs outside of conscious awareness.⁴⁶

Of course, this pattern also holds for civil claims derived from statutes or constitutional provisions. For example, in both statutory and

and 6) as a result of relying on the representation, the plaintiff sustained damages. See BERNARD E. WITKIN, SUMMARY OF CALIFORNIA LAW §§ 676-677, 778-79 (9th ed. 1988).

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

45. Celotex Corp. v. Catrett, 477 U.S. 317 (1986); CHARLES ALAN WRIGHT, ET AL., § 2727 10A (3d ed. 1998).

46. For an excellent treatment of this and related phenomena, and their implications for the law of securities fraud, see Donald C. Langevoort, *Organized Illusions: A Behavioral Theory of Why Corporations Mislead Stock Investors (and Cause Other Social Harms)*, 146 U. PA. L. REV. 101, 139-48 (1997).

constitutional antidiscrimination law, the distinction between disparate treatment theory and disparate impact theory inherently relies on the assumption that the “intent to discriminate” required to establish a disparate treatment claim is psychologically distinct and practically distinguishable from the mindless indifference to harms inflicted on an outgroup by facially neutral policies.

In most cases, it is not merely the evidence placed into the record that establishes or disestablishes an essential element of the particular claim for relief for which it was offered. Rather, in most situations, the essential elements of a litigant’s case are proven (or not) through the presentation of numerous evidentiary facts, which, combined with inferences reasonably drawn from those evidentiary facts, either do or do not establish the existence of each essential element by a preponderance of the evidence.⁴⁷

In the adjudication of summary judgment motions, judges use implicit theories of human behavior to determine, among other things, what inferences would be “reasonable” or “unreasonable” to draw from a particular factual pattern, since only “reasonable” inferences need be drawn in favor of the non-moving party (usually the plaintiff) and against the moving party (usually the defendant).⁴⁸

Assuming that a case survives summary judgment and gets to trial, jurors then need theories of human behavior to know which of various competing “reasonable” inferences should be drawn from the facts adduced in evidence. In most cases, proof is circumstantial; inferences form a kind of logical bridge between the evidentiary facts adduced in evidence and the ultimate facts that must be established or disestablished for one or the other litigant to prevail.

Obviously, in cases involving social interaction, one’s implicit or explicit theories of human and organizational behavior will play a central role in drawing inferences from facts, and in determining whether those inferences are reasonable. Generally speaking, people view particular inferences as relatively more reasonable, and are more likely to draw them, if such inferences comport with their subjective understanding of the common nature of things in the world. For this reason, in any lawsuit involving human social interaction, judgment, or choice, both judge and jury act as “intuitive psychologists,”⁴⁹ bringing to bear on the unfolding case narrative their implicitly held theories of human behavior.

47. See *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574, 587, 592 (1986) (qualifying the well-settled rule that, on summary judgment, all inferences must be drawn in favor of the nonmoving party with the principle that such inferences must be reasonable).

48. *Reeves v. Sanderson Plumbing Prods.*, 530 U.S. 133, 150 (2000).

49. See Ross, *supra* note 33, at 337.

3. Behavioral Theories and Moral Authority in Constitutional Interpretation

In constitutional lawmaking, judges sometimes use psychological theories to justify decisions based on independent normative grounds. In this regard, consider the “separate but equal” principle pronounced in *Plessy v. Ferguson*⁵⁰ and repudiated fifty-six years later in *Brown v. Board of Education*.⁵¹ As described in this Article’s introduction, many commentators criticized the Warren Court’s use of social science evidence in its decision in *Brown*. But for the most part, these critics fail to acknowledge that *Plessy* used social science, too.

The *Plessy* Court’s interpretation of the equal-protection mandate was analytically based on and rhetorically justified by the essentially psychological claim that laws relegating Blacks to separate public accommodations could effectuate equality, not simply as a legal conclusion, but also in terms of the lived experience of the Black citizens whose lives these laws affected. Specifically, Justice Brown wrote for the *Plessy* majority that state enforced separation of the races stamps Blacks with a badge of inferiority only “because the colored race chooses to put that construction upon it.”⁵² This statement asserted an essentially empirical claim—that the social and personal meaning of Jim Crow, and hence its effect on an individual Black person’s psyche, was simply a product of subjective interpretation that could easily be transformed into something psychologically benign through the exercise of individual or collective cognitive free will.

Fifty-six years later, in *Brown*, the Court reversed course on both the law and the psychology of state-sanctioned segregation. Relying in part on empirical research by Dr. Kenneth Clark and other social scientists, the *Brown* Court concluded, “To separate [Black children] from others of similar age and qualifications solely because of their race generates a feeling of inferiority as to their status in the community that may affect their hearts and minds in a way unlikely ever to be undone.”⁵³

50. 163 U.S. 537 (1896).

51. 347 U.S. 483 (1954).

52. Justice Brown stated in *Plessy*:

Laws permitting, and even requiring, [the] separation [of the races] in places where they are liable to be brought into contact do not necessarily imply the inferiority of either race to the other. . . . We consider the underlying fallacy of the plaintiff’s argument to consist in the assumption that the enforced separation of the two races stamps the colored race with a badge of inferiority. If this be so, it is not by reason of anything found in the act, but solely because the colored race chooses to put that construction upon it.

Plessy v. Ferguson, 163 U.S. 537, 544, 551 (1896).

53. *Brown v. Bd. of Educ.*, 347 U.S. 483, 494 (1954). The Court in *Brown* went on to quote a finding by the district court that:

Segregation of white and colored children in public schools has a detrimental effect upon the colored children. The impact is greater when it has the sanction of the law; for the policy of separating the races is usually interpreted as denoting the inferiority of the negro group. A sense of inferiority affects the motivation of a child to learn. Segregation with the sanction of law, therefore, has a tendency to [retard] the educational and mental development of negro

As this passage suggests, the *Brown* Court's interpretation of the equal protection mandate was based on and justified by an empirically testable claim about the psychological effects of segregation on Black children.⁵⁴ The social science research described in *Brown*'s footnote eleven functioned to justify not only the Court's decision on the merits, but also its departure from the principle of *stare decisis*, as Justice Sandra Day O'Connor noted in *Planned Parenthood of Southeastern Pennsylvania v. Casey*.⁵⁵ The logic of *Brown* suggests that advances in empirical social psychology not only allowed but even compelled the Court to repudiate the separate but equal doctrine, on the grounds that the psychology relied upon by the *Plessy* majority had been not only publicly repudiated but also empirically invalidated.

Brown v. Board of Education is, of course, not the only Fourteenth Amendment case in which the Supreme Court invoked social science theories to justify its decision. Almost fifty years earlier, in *Muller v. Oregon*,⁵⁶ the Court upheld an Oregon law restricting the daily number of hours women could work, basing its decision in part on the contents of the original "Brandeis brief," which surveyed medical and social science theories about the effect of long working hours on women's physical and mental health.⁵⁷ More recently, social science theories reported in the amicus contributions of organizations such as the American Psychological Association and the American Sociological Association were used by the Supreme Court majorities and minorities in the Michigan affirmative action cases of *Gratz v. Bollinger*⁵⁸ and *Grutter v. Bollinger*.⁵⁹

children and to deprive them of some of the benefits they would receive in a racial[ly] integrated school system.
Id. at 494.

54. We are aware that the question of whether the Warren Court was, in fact, influenced by the results of Kenneth Clark's doll study and other social science research concerning the effects of segregation on Black children is contested. Compare, e.g., Sanjay Mody, *Brown Footnote Eleven in Historical Context: Social Science and the Supreme Court's Quest for Legitimacy*, 54 STAN. L. REV. 793 (2002) (arguing against the claim) with Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 YALE L.J. 1535 (1998) (arguing in favor of the claim). See generally JUAN WILLIAMS, THURGOOD MARSHALL: AMERICAN REVOLUTIONARY 197-205 (1998) (describing NAACP's decision to submit psychologist Kenneth Clark's "doll study" as evidence of segregation's harmful effect on black children); MARK V. TUSHNET, THE NAACP'S LEGAL STRATEGY AGAINST SEGREGATED EDUCATION, 1925-1950 (1987) (describing early stages of litigation that led to the 1954 decision in *Brown*); HERBERT HILL & JACK GREENBERG, CITIZEN'S GUIDE TO DESEGREGATION: A STUDY OF SOCIAL AND LEGAL CHANGE IN AMERICAN LIFE (1955) (describing in Chapters 7 and 8 the pre-1954 stages of the *Brown* litigation).

55. 505 U.S. 833, 863 (1992).

56. 208 U.S. 412 (1908).

57. For a discussion of the use of social science evidence in *Muller v. Oregon*, 208 U.S. 412 (1908), see John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477, 480-82 (1986).

58. 539 U.S. 244 (2003).

59. 539 U.S. 306 (2003).

In each of these three instances, social science was arguably used to justify the Court's choice of a legal rule based on independent normative grounds. In *Muller*, insights from the medical and social sciences were used to bolster the Court's decision that the Oregon law restricting women's, but not men's, working hours was rational. In *Brown*, social science findings relating to the psychological effects of segregation on Black children were used to legitimate the Court's repudiation of *Plessy v. Ferguson* and the "separate but equal" doctrine it embraced. And, in *Gratz and Grutter*, the Court's minority and majority opinions, respectively, based their qualified defense of preferential forms of affirmative action on a set of essentially empirical claims about the effect of student body diversity on educational quality. Reading any of three of these cases, one could easily be left with the troubling sense that the Court had used the normalizing power of science to airbrush away the sharp edges of a difficult and divisive normative choice.

We return in our conclusion to problems stemming from the misuse of social science in substantive lawmaking. For present purposes, however, this counterpoint is key: those who criticize the use of insights from the empirical social sciences in constitutional lawmaking often fail to recognize that the alternative to empiricism is often bare judicial surmise, posing as common sense, and having a lesser claim to validity than the imperfect science such critics would exclude from the legal analytical process.

4. *The Use of Behavioral Theories in the Shaping of Legal Incentives*

Two of the core insights of the law and economics movement are that people respond to incentives and that law can serve as a powerful tool for structuring those incentives in socially beneficial ways.⁶⁰ It is testimony to the enormous success of this movement that legal rules are increasingly crafted with an eye toward structuring incentives so as to encourage people and organizations to act in particular ways and forbear from acting in others. Indeed, one finds all across the legal landscape the shaping of statutes, regulations, and judicial interpretations of statutes and regulations in ways explicitly designed to encourage desirable behaviors and discourage undesirable ones within a particular compliance community.⁶¹

When judges craft legal rules to shape incentives and thereby further the law's normative goals, they necessarily draw upon theoretical models of individual and organizational motivation and behavior. Sometimes the models on which judges draw are simply assumed to be accurate, even

60. Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CALIF. L. REV. 1051, 1054 (2000).

61. By "compliance community" we mean those individuals and entities that are either responsible for complying with a particular legal regime or entitled to its protection.

though they represent testable theories that have been subjected to scrutiny in the empirical social sciences.

In the antidiscrimination law context, the legal doctrine governing employer liability for sexual harassment by supervisors illustrates how judges sometimes use unexamined, often even unstated, behavioral theories when they craft legal rules with an eye toward structuring individual and organizational incentives. For example, in two cases decided in 1998, *Faragher v. City of Boca Raton*⁶² and *Burlington Industries v. Ellerth*,⁶³ the Supreme Court established an affirmative defense to Title VII claims of hostile work environment harassment by supervisors. The defense permits an employer to defeat an otherwise meritorious hostile environment harassment claim if it can show that it had promulgated antiharassment policies and instituted antiharassment education and grievance procedures, and that the plaintiff had failed to use those procedures early in an escalating sequence of harassing events.⁶⁴

Writing for the *Faragher* majority, Justice Souter acknowledged that the Court's decision to establish the affirmative defense deviated from a long line of cases applying agency principles to hold employers responsible for tortious conduct by their supervisory employees.⁶⁵ Justice Souter justified this departure, however, on utilitarian grounds. By providing employers with incentives to institute antiharassment policies, education programs, and grievance procedures, and by encouraging employees to complain early in an escalating sequence of harassing events, the new affirmative defense would serve Title VII's "primary objective," which Justice Souter described as "not to provide redress but to avoid harm."⁶⁶ As various commentators have noted,⁶⁷ the Court's reasoning in *Faragher* reflects and

62. *Faragher v. City of Boca Raton*, 524 U.S. 775 (1998).

63. *Burlington Indus. v. Ellerth*, 524 U.S. 742 (1998).

64. *Faragher*, 524 U.S. at 807-08; *Ellerth*, 524 U.S. at 764-65.

65. *Faragher*, 524 U.S. at 804.

66. *Id.* at 806.

67. See, e.g., THERESA M. BEINER, *GENDER MYTHS v. WORKING REALITIES: USING SOCIAL SCIENCE TO REFORMULATE SEXUAL HARASSMENT LAW* (2005) (reviewing social science evidence regarding patterns of women's response to unwanted sexualized workplace conduct); Anne Lawton, *Tipping the Scales of Justice in Sexual Harassment Law*, 27 OHIO N.U. L. REV. 517 (2001) (examining social-science evidence bearing on the utility of internal grievance procedures in preventing or remedying sexual harassment); Linda Hamilton Krieger, *Employer Liability for Sexual Harassment—Normative, Descriptive, and Doctrinal Interactions: A Reply to Professors Beiner and Bisom-Rapp*, 24 U. ARK. LITTLE ROCK L. REV. 169 (2001) [hereinafter Krieger, *Employer Liability for Sexual Harassment*] (using social-science evidence to question the assumption implicit in Supreme Court harassment doctrine that "reasonable" women who are exposed to unwelcome sexualized conduct in the workplace will file internal complaints with their employers early in an escalating sequence of harassing incidents); Susan Bisom-Rapp, *Fixing Watches with Sledgehammers: The Questionable Embrace of Employee Sexual Harassment Training by the Legal Profession*, 24 U. ARK. LITTLE ROCK L. REV. 147 (2001) (commenting on the dearth of social science evidence supporting the Supreme Court's assumptions that antiharassment policies, training programs, and grievance procedures actually prevent harassment from occurring).

relies upon a particular set of unstated and unexamined, but empirically testable, assumptions. First, the Court assumes that by promulgating policies against harassment and by establishing antiharassment education and grievance procedures, employers will prevent harassment from occurring, or will, at the very least, significantly decrease its incidence. Second, the Court's analysis rests on the premise that employees who are subjected to unwelcome sexualized workplace conduct will use an employer-run grievance procedure if one is provided, and that they will do so before the conduct becomes egregious.

However, as numerous commentators have documented, virtually no empirical support exists for the proposition that antiharassment policies, training programs, or internal grievance procedures actually reduce the amount of unwanted sexualized conduct in the workplace.⁶⁸ Moreover, a large amount of empirical evidence indicates that reporting through existing internal grievance mechanisms is an extremely rare response to unwanted sexualized workplace conduct, and that women generally have quite reasonable justifications for their decision not to report.⁶⁹ In other words, in their efforts to use the *Faragher/Ellerth* affirmative defense to structure incentives and ultimately prevent harassment from occurring, the justices appear to have presumed a behavioral world that does not in fact exist.

If sexual harassment doctrine is premised on the essentially empirical claim that antiharassment policies, training, and grievance procedures reduce harassment, or that "reasonable" women will use those procedures early in an escalating sequence of harassing events, and if these claims are empirically false, the justices' attempt to deploy the *Faragher/Ellerth* affirmative defense to prevent harassment from occurring will fail to achieve its policy objectives. Rather than preventing harassment from occurring, the affirmative defense will simply operate to defeat otherwise meritorious harassment claims.

It is significant to note how, in *Faragher* and *Ellerth*, the Court simply takes as given that employer antiharassment policies and procedures will reduce the incidence and severity of harassment. The Court's commitment to this "common sense" view deepened in 1999, when in *Kolstad v. American Dental Association*,⁷⁰ the Court established an affirmative

68. BEINER, *supra* note 67; Joanna L. Grossman, *The Culture of Compliance: The Final Triumph of Form Over Substance in Sexual Harassment Law*, 26 HARV. WOMEN'S L.J. 3 (2003); Bisom-Rapp, *Fixing Watches with Sledgehammers*, *supra* note 67; Bisom-Rapp, *An Ounce of Prevention is a Poor Substitute for a Pound of Cure*, *supra* note 21; Theresa M. Beiner, *Sex, Science and Social Knowledge: The Implications of Social Science Research on Imputing Liability to Employers for Sexual Harassment*, 7 WM. & MARY J. WOMEN & L. 273 (2000); Krieger, *Employer Liability for Sexual Harassment*, *supra* note 67.

69. See sources cited *supra* note 68.

70. 527 U.S. 526 (1999).

defense against Title VII claims for punitive damages similar to the defense it established in the sexual harassment context. Under this affirmative defense, employers can avoid liability for punitive damages by promulgating antidiscrimination policies and attempting to educate their employees about Title VII's prescriptions.⁷¹ Writing for the *Kolstad* majority, Justice Sandra Day O'Connor echoed Justice Souter's statement in *Faragher* that Title VII's primary objective was not to provide redress, but to prevent harm.⁷² For that reason, she continued, the purposes underlying Title VII are advanced when employers are encouraged to adopt formal antidiscrimination policies and inform their employees about antidiscrimination laws. The affirmative defense will provide incentives for employers to adopt antidiscrimination policies and training programs, and this, the Court reasoned, will reduce the incidence of discrimination, thus furthering Title VII's goals.

Throughout the Court's analysis in *Kolstad*, as in *Faragher* and *Ellerth*, the Court simply assumes that antidiscrimination policies and training programs reduce the incidence of discrimination. There is virtually no empirical evidence, however, to support this claim. In fact, there is some evidence suggesting exactly the opposite.⁷³

Of course, apart from judicial attempts to engineer incentives, legal rules can be devised to serve other functions about which empiricism has little if anything to say. So, for example, the Court could have premised the *Faragher/Ellerth* defense on the notion that it would be unfair to impose liability on a corporation that is trying, however imperfectly, to ensure that its employees comply with the ambiguous and often conflicting set of legal and practical considerations that sexual-harassment law implicates. Alternately, the Court might have based the defense on the grounds that, as a matter of judicial economy, if an employer provides an internal grievance mechanism for redressing harassment claims and if an allegedly aggrieved employee has failed to use it, she should not be allowed to seek redress from the already overburdened federal court system. In other words, legal doctrines can be, and often are, premised on normative principles about which, in the final analysis, empirical research has little to say.

For example, were it empirically established that state-sanctioned segregation does *not* injure Black children's self-esteem, would we want to reverse *Brown*? How many supporters of preferential forms of affirmative action would change their views as to its constitutionality if it were established that student body diversity had no discernable effects on educational

71. See *id.* at 545-46.

72. *Id.* at 545.

73. See, e.g., Bisom-Rapp, *Fixing Watches with Sledgehammers*, *supra* note 67 (reviewing available research record and concluding that little if any evidence supports the proposition that antidiscrimination policies and training programs actually reduce the incidence of discrimination or harassment).

outcomes? As these questions illustrate, empiricism is no substitute for normative analysis.

However, empiricism is indispensable to sound consequential analysis. Obviously, a legal rule can effectively shape conduct through the structuring of individual or organizational incentives only if that rule is built upon accurate models of individual and organizational behavior and choice. When judges premise legal doctrines on the notion that a particular legal rule will structure individual and organizational incentives and thereby shape behavior in a particular way, they need to get the social science right. Although far from infallible, psychological science often serves as a better source of such models than “common sense” or other forms of *a priori* theorizing.

B. Behavioral Theories in Cognitive Social Psychology

What then does it mean to “get the science right”? How do social-scientific communities go about evaluating the theories that structure their investigations, and what, if anything, can jurisprudence learn from science about its own relationship to the psychological theories it generates and employs?

When creating or updating behavioral theories, psychological science employs a wholly empirical standard of proof. Normative analysis has no role to play in determining whether a behavioral theory is sound. In psychological science, investigations proceed from hypotheses, which state the expected relationship between two or more variables.⁷⁴ Usually, causality is stated (A increases B) or strongly implied (A and B go together, and A is temporally prior). Hypotheses most often derive from theories supported by basic research (i.e., research concerned with fundamental principles rather than with the application of those fundamental principles to a specific applied problem).⁷⁵ Scientists share criteria for what makes good theory. These criteria include causality, coherence, parsimony, and falsifiability.⁷⁶

Questions about causation occupy most scientific research, and social science is no exception. Well-established empirical standards for causal inference require that: (a) two variables be associated (co-vary; change together over time, occasions, or persons); (b) the proposed cause temporally precede the proposed effect; (c) the relationship is not spurious (i.e., in reality jointly caused by a third variable); and (d) a rationale be argued.⁷⁷ The

74. SUSAN T. FISKE, *SOCIAL BEINGS: A CORE MOTIVES APPROACH TO SOCIAL PSYCHOLOGY* 40 (2004).

75. *Id.* at 38.

76. *Id.* at 38-39.

77. *Id.* at 60-61.

following discussion will briefly elaborate on each one of these requirements.

To meet these standards, an individual conducting an experiment—the gold standard for inferring causality—would first show temporal priority by manipulating experimental conditions and measuring the result. Next, she would establish association by measuring the result at different values of the putative cause. She would also demonstrate nonspuriousness by having randomly assigned subjects to different conditions and having controlled for extraneous variables. Finally, theory would provide the rationale by explaining mechanisms linked to a network of other theories and observations. Without theory, empirical observations would merely collect a series of isolated facts without larger meaning.

In addition to inferring causality, a good theory possesses coherence. It must not contradict itself. It must reflect internal logic, and it must tell a plausible story.

Good theories are also parsimonious. So long as accuracy is not compromised, explaining a phenomenon with a few measurable variables is preferable to explaining it with many. Unfortunately, nature does not always cooperate. Elegance makes theories more esthetically pleasing, and more testable, but it may not make them more accurate. Neoclassical economics provides a salient example. Economic theories based on the rational-actor assumption are elegant and parsimonious, but as research in behavioral economics has shown, they do not accurately predict much of human economic behavior.⁷⁸

Finally, and perhaps most importantly, to be legitimate a social scientific theory must be falsifiable. The central criterion for falsification is that observations, made under suitably controlled conditions, can contradict the theory being tested. Enough of these observations, replicated over time and investigator, topple the theory, and new contenders emerge.⁷⁹

The evidence for and against any previously accepted social-scientific theory is subject to rigorous peer review. As a rule, scientists are conservative and resist new approaches, holding their colleagues to very high standards. New theoretical models that survive this process are then themselves subjected to attempts at falsification.

Thus, theories in psychological science are inherently destined to evolve. Current theories explain the body of observations to date, and new variants incorporate the older observations as well as new ones. The body of data remains constant, though its interpretation shifts as it is viewed through the lenses of evolving theories. That psychological theories change does not make them arbitrary; they must withstand scrutiny under the best methods and data available at the time.

78. See generally *ADVANCES IN BEHAVIORAL ECONOMICS* (Colin F. Camerer et al. eds., 2004).

79. KUHN, *supra* note 3.

C. *Quality Control:*
Incorporating Psychological Theories into Legal Doctrines

Given the relative impermanence of psychological theories, what standards should govern their incorporation into law? This question, or at least one much like it, occupies an entire field of law and scholarship on the admissibility of expert testimony, and we do not intend to review or add directly to that literature here. A set of basic principles, however, bear mention.

At a minimum, as suggested by the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals*,⁸⁰ the law should not incorporate a psychological theory into one of its many doctrines unless it has been empirically tested, has been subject to peer review and publication, has garnered widespread acceptance within the relevant scientific community, and, where applicable, has a known and acceptable error rate.⁸¹ In connection with factual adjudication, a trial court, applying *Daubert*, plays a gate-keeping role; it screens proffered scientific evidence to determine whether it satisfies applicable scientific standards, and whether it would be useful to the trier of fact.

Social science models must clear an additional hurdle before use in adjudication. Psychological science is basic research, often conducted in the laboratory. Meanwhile, legal disputes concern events in the real world. As such, questions of "external" or "field" validity inevitably arise.⁸² The challenge often takes this form: does the experimental result pertain only to the rarefied, highly controlled, artificial context of the laboratory, or does it also accurately represent how things work in the real world?

A theory might have laboratory but not field validity for a number of reasons. First, by controlling for extraneous factors, a well-designed experiment might amplify an effect that outside the laboratory is dominated by the very factors being controlled. Second, a laboratory finding may be statistically significant at the 0.01 or 0.05 level even when the actual effect size is relatively small.⁸³ Third, the dependent variable measured in a

80. 509 U.S. 579 (1993).

81. *Id.* at 593-94. *Daubert* frames the critical issue as one involving the relevant theory's "reliability." The word choice is somewhat unfortunate. Within scientific communities, "reliability" refers to the consistency with which a researcher can make a particular experimental observation. A theory's accuracy or soundness, on the other hand, goes to the issue of its validity. Validity was the concern in *Daubert*.

82. FISKE, *supra* note 74, at 68.

83. An observed difference or relationship among variables is said to be "statistically significant" if the odds of that difference or relationship occurring by chance are less than 5% (in which case it is significant at the .05 level) or less than 1% (in which case it is significant at the .01 level). The term "effect size" is used to indicate the practical extent of the difference on some characteristic of interest between the experimental and control groups in a controlled experiment. MARY B. HARRIS, *BASIC STATISTICS FOR BEHAVIORAL SCIENCE RESEARCH* 282, 537 (2nd ed. 1995).

laboratory demonstration of a particular effect may differ from the actual real world behavior it was designed to simulate.

Using primary research in psychological science implicates all of these issues, but tools exist to reckon with them. Effect sizes can be measured and expressed as a standardized statistical unit.⁸⁴ Researchers can use representative-sample survey and field observations to assess the field validity of a theory validated in the laboratory environment.

Individuals who question the field validity of laboratory demonstrations often base their criticisms on a questionable assumption. These individuals assume that by removing most sources of noise, the laboratory setting increases rather than decreases the strength of the studied phenomenon's signal. This assumption flatters social science by likening it to the field of pure physics, comparing a reduced-noise psychology experiment with a frictionless vacuum. But people are not particles, and the noise in a social system tends to conform to other noise in the system. Fortunately, through the use of meta-analysis, which allows the quantitative summary of results across multiple studies, social scientists can sometimes identify field effects that are *larger* than their laboratory-bound counterparts.

Several such meta-analyses have been done on phenomena pertaining to prejudice and discrimination. Stronger field than laboratory effects have demonstrated the real world generality of favoring members of one's ingroup,⁸⁵ remembering stereotypic more than counterstereotypic information,⁸⁶ judging the outgroup as more homogeneous than the ingroup,⁸⁷ and liking people one sees frequently.⁸⁸

In short, psychological science provides effective mechanisms not only for testing the laboratory validity of psychological theories, but also for determining whether those theories accurately represent effects that occur with meaningful frequency in the real world. And even where only a few studies demonstrating field validity exist, it is hard to understand why judge's *a priori* psychological intuitions, which have no demonstrated validity at all, should be preferred to psychological theories validated by

84. One statistical technique used to express effect size is the Cohen's *d*, which equals the difference between the means of the experimental and control groups divided by the standard deviation of the control group. See generally, Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences* (1988) (describing the importance of reporting effect sizes in social science research and proposing a specific test, now known as "Cohen's *d*," for expressing effect size as a standard measure).

85. Brian Mullen, Rupert Brown & Colleen Smith, *Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration*, 22 EUR. J. OF SOC. PSYCHOL. 103 (1992).

86. Charles Stangor & David McMillan, *Memory for Expectancy-Congruent and Expectancy-Incongruent Information: A Review of the Social and Social Developmental Literatures*, 111 PSYCHOL. BULL. 42 (1992).

87. Thomas M. Ostrom & Constantine Sedikides, *Out-Group Homogeneity Effects in Natural and Minimal Groups*, 112 PSYCHOL. BULL. 536 (1992).

88. See Robert F. Bornstein, *Exposure and Affect: Overview and Meta-Analysis of Research, 1968-1987*, 106 PSYCHOL. BULL. 265 (1989) (noting the extensive research that shows that repeated exposure leads to an "increase in positive affect" to a particular stimulus).

solid, laboratory-based science. As we noted above, *Daubert* stands for the proposition that adjudicative facts should not be based on a psychological theory unless that theory has been empirically tested, subjected to peer review and publication, has garnered widespread acceptance within the relevant scientific community, and, where applicable, has a known and acceptable error rate. But, as we have shown and will demonstrate in greater detail in Part II, in elaborating legal doctrines and in applying them in legal reasoning, judges routinely articulate and apply intuitive psychological theories that satisfy none of these normative criteria. And once incorporated into legal doctrine, these lay psychological theories can be quite difficult to modify or uproot.

D. Modifying Behavioral Theories in Law:

When is Theory Change Needed, and Why Is It So Hard to Effect?

As we have shown, as judges develop, interpret, explain, and apply legal rules, they necessarily construct and draw upon models of human behavior they assume to be empirically sound. They do this in a number of ways: by incorporating certain essential elements of proof into the claims for relief they recognize as legally cognizable; by the language they use to explain why, in any particular case, those elements have or have not been established; by their crafting of jury instructions; and through their definition of ambiguous statutory terms. Judges make assumptions about human behavior in their evidentiary rulings and in their written opinions granting or denying summary judgment or judgment as a matter of law. Sometimes records of judicial reasoning explicitly posit these theories of human behavior, and sometimes they do not. Either way, because law's legitimacy is based on the appearance of stability, continuity, and fidelity to precedent, the judiciary faces high costs when it questions or changes the implicit psychology behind established legal doctrines. The kind of course correction taken in *Brown v. Board of Education* can happen only so often if the judiciary is to preserve rule-of-law notions and popular legitimacy.

Litigants, like judges, have little incentive to question the inferential architecture undergirding the substantive legal theories on which their claims or defenses are premised. Although civil litigation is in many ways highly technical, at the end of the day, lawsuits tell stories. Because judicial opinions incorporate popular, taken-for-granted assumptions about the common nature of things, they function as a society's core stories; they offer an interpretation of experience and provide the participants of future lawsuits a narrative comprising a set of easily recognized plots, symbols, themes, and characters.⁸⁹

89. See generally AMSTERDAM & BRUNER, *supra* note 42 (describing the roles of psychological processes including categorization, narrative pattern recognition, and rhetorical cultural resonance in legal advocacy and judicial decision making).

Future litigators fall hostage to these narratives. As the narratives become paradigms—as they become entrenched in legal doctrine—litigators have an irresistible incentive to construct from the facts of their client’s particular case a story that closely resembles them. Challenging the law’s implicit, phenomenological metatheories is a good way for a lawyer to lose her client’s case. In short, whether one is a judge or a litigant, legal institutions structure incentives in ways that protect the implicit theories of human behavior underpinning established legal doctrines from challenge or change.

On the other hand, as the last section described, empirical psychology has a very different relationship to its theories of human behavior. Empirical social science progresses by testing, problematizing, refining, supplementing, and sometimes disconfirming, established theories of human perception, judgment, behavior, and choice. Social psychology is interesting because it produces vivid, replicable demonstrations of how human beings tend to behave in particular contexts. These demonstrations are useful because they often sharply contradict our “common sense” theories of human behavior, including those embedded in law.

Empirical social psychologists, unlike judges and litigants, have powerful incentives to challenge the established theoretical models that constitute their discipline. In a very real sense, empirical psychology develops theories of human behavior as scaffolding for further investigation, which, if successful, conduces over time to the theory’s modification. In empirical psychology, as in all science, theory emendation represents success. In jurisprudence, on the other hand, theory change, unaccompanied by legislation or constitutional amendment, threatens subjective conceptions of judicial legitimacy. For this reason, once a particular model of human behavior becomes embedded in legal doctrine, judges may go to great lengths to avoid having to modify it. As a result, in terms of their accuracy and comprehensiveness, empirically tested psychological theories tend to get ahead of their judicial counterparts.

Numerous examples of this lag between psychological and jurisprudential theories of the same phenomenon come readily to mind. Together, they reveal the critical need for the law to find more effective mechanisms for examining, and where indicated, changing, the behavioral theories it employs. Eyewitness identification testimony, for example, has retained its judicial imprimatur of singular probative value long after empirical social scientists vividly demonstrated its frightful unreliability.⁹⁰ Cases involving repressed memories continue to circumvent otherwise applicable statutes of limitations, despite research by Elizabeth Loftus and others casting the syndrome’s very existence as a genuine phenomenon into serious

90. See generally ELIZABETH F. LOFTUS, *EYEWITNESS TESTIMONY* (2d ed. 1996) (summarizing the body of work on the unreliability of eyewitness testimony).

doubt.⁹¹ And courts continue to admit as evidence expert testimony purporting to predict future violence in convicted murderers⁹² in death penalty cases, despite the work of psychologists like John Monahan, Robert Menzies, and others, who have demonstrated that in most contexts such predictions have little, if any, validity.⁹³

Where social science does influence legal doctrine, the lag-time between scientific and legal acceptance appears to be at least five years. In one respect this lag serves a useful purpose; it allows scientific consensus to solidify, and it inhibits the overeager adoption of fads and flukes. In another respect, however, the five-year lag poses a problem if the alternative is not older science but flawed common sense theories of behavior. Unvalidated intuitions are often a poor substitute for science, even five-year-old science.

Nowhere, we suggest, have the intuitive psychological theories employed by judges in the elaboration of legal doctrine and the adjudication of individual cases fallen further behind their empirically vetted social science counterparts than in the context of antidiscrimination law. In virtually every way that judges presiding over civil rights cases employ implicit psychological theories—in the construction of legally cognizable claims for relief; the schematic abstraction and description of those claims' constituent elements in judicial opinions; the analysis of factual records in connection with summary adjudication motions; the analysis of motions for judgment as a matter of law; the crafting of jury instruction; and substantial-evidence review on appeal—antidiscrimination doctrine lags far behind the psychological science of intergroup bias.

In other words, current judicial models of what discrimination is, what causes it to occur, and how one should go about determining whether it happened in any given case, now diverge in significant ways from the models of intergroup bias validated through empirical inquiry in the social sciences. Furthermore, although from time to time apparent doctrinal inroads have been made, antidiscrimination jurisprudence has in many respects proven stubbornly refractory to the conceptual adjustments the

91. See generally ELIZABETH LOFTUS & KATHERINE KETCHAM, *THE MYTH OF REPRESSED MEMORY* (1994) (questioning the scientific validity of repressed memories); Timothy D. Wilson & Elizabeth W. Dunn, *Self-Knowledge: Its Limits, Value, and Potential for Improvement*, 55 ANN. REV. PSYCHOL. 493, 497-98 (2004) (reviewing more current research on memory suppression).

92. This principle derives from the U.S. Supreme Court's decision in *Barefoot v. Estelle*, 463 U.S. 880, 901-03 (1983), which is still good law.

93. See generally John Monahan, *The Scientific Status of Research on Clinical and Actuarial Predictions of Violence*, in DAVID L. FAIGMAN ET AL., *MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY* 423 (2d ed. 2002); Robert Menzies et al., *The Dimensions of Dangerousness Revisited: Assessing Forensic Predictions About Violence*, 18 LAW & HUM. BEHAV. 1 (1994); David Faust & Jay Ziskin, *The Expert Witness in Psychology and Psychiatry*, 241 SCI. 31 (1988).

science of implicit bias suggests it needs. We now turn our attention to that subject.

II

A BEHAVIORAL REALIST CRITIQUE OF DISPARATE TREATMENT DOCTRINE

The past thirty years of empirical research in social psychology have yielded four fundamental insights of central importance to antidiscrimination law and to theories of proof in discrimination cases. The first concerns the relative roles of controlled and automatic processes in human judgment and choice. Far more appears to go on in the cognitive background, beyond the perceiver's conscious attentional focus, than naïve theories of human behavior suggest. In other words, much of what we do is "mindless,"⁹⁴ proceeding on automatic pilot.⁹⁵

The second principle concerns the importance of subjective construal in shaping human judgment and behavior. Recent research has shown that raw perceptual input does not, on its own, shape impression formation, memory, or eventual choice. Rather, perceptual input is rendered meaningful only through a process of subjective construal.⁹⁶ Construal is subject to a variety of distortions stemming from implicit schematic expectancies, including but not limited to social stereotypes. This biased construal is not unique to intergroup perception. Social perception and judgment more generally are subject to numerous biases, which often impair people's ability to accurately predict the future, attribute causation, and otherwise understand the social world in which they act, judge, and choose. Frequently, as the first principle suggests, these biases may operate outside of the perceiver's conscious, attentional focus. Often, they can be controlled only through the application of effortful, deliberate, cognitive "correction."⁹⁷

Social psychology's third important principle taps into the importance of belonging⁹⁸ to one's primary social groups. People demonstrably conform to their social surroundings far more than they think they would.⁹⁹

94. ELLEN J. LANGER, *MINDFULNESS* (1989).

95. John A. Bargh, *The Automaticity of Everyday Life*, in 10 *ADVANCES IN SOCIAL COGNITION* 1 (Robert S. Wyer, Jr. ed., 1997).

96. See LEE ROSS & RICHARD E. NISBETT, *THE PERSON AND THE SITUATION: PERSPECTIVES OF SOCIAL PSYCHOLOGY* 59-89 (1991); Jerome S. Bruner, *On Perceptual Readiness*, 64 *PSYCHOL. REV.* 123, 123-27 (1957); Lee Ross & Donna Shestowsky, *Contemporary Psychology's Challenges to Legal Theory and Practice*, 97 *Nw. U. L. REV.* 1081, 1088 (2003).

97. See Timothy D. Wilson & Nancy Brekke, *Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations*, 116 *PSYCHOL. BULL.* 117, 130 (1994). See generally, SUSAN T. FISKE & SHELLEY E. TAYLOR, *SOCIAL COGNITION* (2d Edition, 1991) (describing social construal and correction processes in social perception).

98. Roy F. Baumeister & Mark R. Leary, *The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation*, 117 *PSYCHOL. BULL.* 497, 497 (1995).

99. See generally Solomon E. Asch, *Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority*, in 70 *PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED* 1 (Herbert S. Conrad ed., 1956).

This finding holds across cultures and over time.¹⁰⁰ The importance of understanding people as social beings¹⁰¹ permeates all of social psychology. This follows from the fact that people have never survived well outside of primary groups. Due to this core social motive to get along with and feel accepted by one's group, most people attune their behavior to the thinking and feelings of those around them more than introspection would admit. Social context has an enormous impact on individual behavior.

This leads into the fourth key principle emerging from empirical social psychology—the surprising power of situations, as opposed to assumedly stable, dispositional traits or tastes, in shaping judgment, behavior, and choice. Empirical investigations in social psychology have shown that by varying even minor aspects of the situation in which people are asked to judge, act, or choose, one can elicit completely different patterns of behavior, even if nothing about the actor's self-interest has been changed. Situational variables, as contrasted with stable dispositional traits or "tastes," have far more power in shaping preferences and in driving social behavior than most people assume.¹⁰²

These four principles—the important role played by the cognitive unconscious in social perception and judgment; the centrality of subjective construal; the motivational power of belonging; and the power of situations—function together as the weight-bearing pillars of modern social psychology. Significantly, they all directly challenge the "intuitive psychology" underpinning much of currently accepted federal antidiscrimination doctrine. In numerous ways, antidiscrimination law reflects and reifies a common-sense theory of social perception and judgment that attributes disparate treatment discrimination to the deliberate, conscious, and intentional actions of invidiously motivated actors. These actors purportedly know when they act in a discriminatory manner, do so consistently across situations, and then dissemble about the real reasons for their decisions when challenged.¹⁰³

In the remainder of this Part, we map the four pillars of modern social psychology onto two nettlesome issues concerning the Title VII disparate treatment doctrine. In Part II.A, we examine a principle known as the honest belief rule and demonstrate that it derives from an intuitive model of social perception and judgment that has been significantly discredited by empirical social science research. This research, which implicates the above-described principles of automaticity and subjective construal, undermines many of the descriptive assumptions on which the honest belief

100. Rod Bond & Peter B. Smith, *Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task*, 119 PSYCHOL. BULL. 111 (1996).

101. See generally FISKE, *supra* note 74 (reviewing psychological research and theory relating to the drive for social belonging).

102. Ross, *supra* note 33, at 337-84; ROSS & NISBETT, *supra* note 96.

103. Krieger, *The Content of Our Categories*, *supra* note 6, at 1185.

rule is premised. Next, in Part II.B, we scrutinize a doctrine known as the same-actor inference, and argue that it is based on a naïve psychology of cross-situational consistency disconfirmed by research reflecting the principle of behavioral situationism and the need for social belonging. Through these analyses, we illustrate both the aims and the methods of behavioral realism in employment discrimination law.

*A. Discriminatory Motivation, Implicit Cognition,
and the Honest Belief Rule*

Section 703(a)(1) of Title VII of the Civil Rights Act of 1964, the statute under which most federal employment discrimination cases arise, provides in pertinent part that it is an unlawful employment practice “to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual’s race, color, sex, national origin, or religion”¹⁰⁴

This statute makes it unlawful to discriminate, yet it never defines exactly what discrimination is. Congress, in essence, left to the courts the task of assigning meaning to the term “discrimination” through a process of judicial elaboration in case law. Given that race, sex, national origin and other forms of discrimination implicate social perception, judgment, and choice, this process of judicial elaboration necessarily relies on descriptive models of how people go about making sense of themselves and others in the social world. For this reason, judicial understandings of the nature of the psychological processes involved in discrimination play a legally constitutive role—that is, they determine the ultimate legal meaning of discrimination.

Despite years of judicial elaboration, the meaning of discrimination is still a work in progress. Following the Supreme Court’s 2003 decision in *Desert Palace, Inc. v. Costa*,¹⁰⁵ an individual seeking to prove a discrimination case under Title VII’s disparate treatment theory must establish three things: 1) that s/he is a member of a Title VII protected class; 2) that s/he was subjected to some negative employment action (e.g., a failure to hire or promote, or a discharge from employment); and 3) that her/his protected group status was a “motivating factor” in the challenged decision.¹⁰⁶ The critical question of course, left for the courts’ future explication, is what it means for one’s protected-group status to be a “motivating factor” in an employment decision. In Part III below, we offer a constructive account,

104. 42 U.S.C. § 2000e-2(a)(1) (2000).

105. 539 U.S. 90 (2003).

106. *Id.* at 101. See generally Jeffrey A. Van Detta, “Le Roi est Mort; Vive Le Roi!”: *An Essay on the Quiet Demise of McDonnell Douglas and the Transformation of Every Title VII Case After Desert Palace, Inc. v. Costa Into a “Mixed Motives” Case*, 52 *DRAKE L. REV.* 71 (2003) (explaining the import of *Costa* for the elements of proof in a Title VII disparate treatment case).

informed both by behavioral realism and traditional methods of statutory interpretation, of how, in our view, the phrase “motivating factor” should be understood for Title VII purposes. In this Part, on the other hand, we seek to show that many federal courts are according the phrase “motivating factor” an unduly crabbed meaning, in part because they are basing their legal analysis on faulty psychological models of intergroup perception and judgment. This, we will argue, leaves many instances of disparate treatment discrimination doctrinally unintelligible and therefore judicially irreparable.

The doctrinal roots of this problem are reflected, ironically, in the plurality opinion in *Price Waterhouse v. Hopkins*,¹⁰⁷ a case we discussed in Part I and in which one of the authors of this article testified as an expert witness on the psychological science of cognitive bias,¹⁰⁸ and which was widely considered a major civil rights victory. Consider again how Justice Brennan, writing for the *Price Waterhouse* plurality, described what it means for gender bias to “motivate” a challenged employment decision:

In saying that gender played a motivating part in an employment decision, we mean that, if we asked the employer at the moment of the decision what its reasons were and if we received a truthful response, one of those reasons would be that the applicant or employee was a woman.¹⁰⁹

This statement reflects an assumption that, when disparate treatment discrimination occurs, the discriminator is, at the moment a decision is made, consciously aware that he or she is discriminating. Gilbert Ryle refers to this assumption as “the official theory” of transparent mental process and describes it as follows:

[A] person has direct knowledge of the best imaginable kind of the workings of his own mind. Mental states and processes are (or are normally) conscious states and processes, and the consciousness which irradiates them can engender no illusions and leaves the door open for no doubts. A person’s present thinkings, feelings and willings, his perceivings, rememberings and imaginings are intrinsically “phosphorescent”; their existence and their nature are inevitably betrayed to their owner.¹¹⁰

Research on implicit social cognition has substantially discredited the “official theory” Ryle described. Much of human mental process, including those processes mediating interpersonal perception and judgment, occur in

107. 490 U.S. 228, 250 (1989).

108. Dr. Fiske testified as an expert witness for the plaintiff, Ann Hopkins. For a description of that testimony and the circuit and Supreme Court’s reactions to it, see generally Martha Chamallas, *Listening to Dr. Fiske: The Easy Case of Price Waterhouse v. Hopkins*, 15 VT. L. REV. 89 (1990).

109. *Price Waterhouse*, 490 U.S. at 250.

110. GILBERT RYLE, *THE CONCEPT OF MIND* 13-14 (1949).

“the cognitive unconscious,”¹¹¹ outside of the perceiver’s mindful attentional focus. In other words, actors often do not realize that they have formed biased judgments of others. If you ask an employer at the moment of the decision what his reasons for making a decision were, he might well not be aware that one of the reasons was that the applicant or employee was a woman, even if her sex did, in fact, influence his judgment.

Perhaps no single experiment better reveals how implicit stereotypes can function as “Trojan Horses,”¹¹² distorting the social judgment of even well-intentioned people, than a well-known study by Princeton psychologists John Darley and Paget Gross.¹¹³ In an investigation of what they called the cognitive confirmation effect, Darley and Gross had subjects view one or two videotapes portraying a school-age child named “Hannah.” By manipulating the environment in which Hannah appeared, Darley and Gross led subjects to believe that her family was either disadvantaged or affluent.¹¹⁴ Immediately after watching one or the other videotape, a subgroup of subjects in each income condition was asked to predict whether Hannah would perform academically “at her grade level,” “above grade level,” or “below grade level.”¹¹⁵ Reluctantly, and often under protest at being asked to base a judgment on stereotypes, subjects in both conditions tended to predict that Hannah would perform “at her grade level.”¹¹⁶

Before being asked to predict Hannah’s ability level, a second subgroup in each income condition was shown an additional videotape. This videotape, identical across both conditions, depicted Hannah responding verbally to twenty-five achievement test problems. Subjects were told that the test included “easy, moderate, and difficult problems.”¹¹⁷ Hannah’s academic performance in this second video was deliberately designed to be ambiguous. After watching the second video, subjects in the low-income and high-income conditions were asked to evaluate Hannah’s performance. At this point, significant differences in the two income conditions emerged. Subjects in the “affluent Hannah” condition rated her academic ability significantly higher than subjects in the “disadvantaged Hannah” condition. Open-ended comments revealed differences in perception as well. For

111. John F. Kihlstrom, *The Cognitive Unconscious*, 237 *Sci.* 1445, 1445 (1987).

112. Jerry Kang, *Trojan Horses of Race*, 118 *HARV. L. REV.* 1489 (2005).

113. John M. Darley & Paget H. Gross, *A Hypothesis-Confirming Bias in Labeling Effects*, 44 *J. PERSONALITY & SOC. PSYCHOL.* 20 (1983).

114. In the “low income” condition, the video showed Hannah playing in a stark, fenced-in, asphalt school yard in a run-down urban neighborhood. Subjects in this condition were told that Hannah’s parents had only a high school education, that her father was employed as a meat packer, and that her mother was a seamstress who worked at home. In the “high income” condition, subjects watched Hannah play in a tree-lined park in an upper middle class neighborhood. Her school was depicted as a large, attractive modern structure, with adjacent playing fields and a shaded playground. Her father was described as an attorney, her mother as a freelance writer. *Id.* at 23.

115. *Id.* at 20.

116. *Id.*

117. *Id.* at 23.

example, subjects frequently described the low-income Hannah as having “difficulty accepting new information,”¹¹⁸ while they described high-income Hannah as demonstrating the “ability to apply what she knows to unfamiliar problems.”¹¹⁹

For our purposes, the key question is, if left uncorrected, would the type of biased construal to which low-income Hannah was subjected in the Paget and Gross study constitute a “motivating factor” within the meaning of Title VII? Surprisingly, the Supreme Court has never squarely addressed this precise question.

Federal disparate treatment case law does make clear that Title VII prohibits employers from requiring that applicants or employees conform to consciously endorsed, prescriptive stereotypes about what social positions members of particular groups (e.g., men versus women, Whites versus Blacks) should occupy or how members of particular social groups should behave.¹²⁰ The case law also plainly establishes that Title VII is violated when employers deliberately use descriptive stereotypes about the characteristics associated with various demographic groups as conscious proxies for traits relevant to job performance.¹²¹

Taken as a whole, these cases reflect a particular conception of how consciously endorsed discriminatory beliefs express themselves as discriminatory behavior that violates Title VII. Under this conception, stereotypes are synonymous with consciously held beliefs, either about how members of distinct social categories should behave (normative beliefs), or about how members of distinct social categories are likely to behave (descriptive beliefs). Encountering a member of one of these groups, a biased decision maker maps the group member into a social category. That category activates the consciously endorsed descriptive and normative beliefs associated with the category. These in turn generate in the bias-holder a conscious, belief-consistent behavioral intention toward the target, which, unless actively overridden, directs the bias-holder’s choices about and behavior toward the stereotyped target.

But social cognition research demonstrates that stereotypes do not function only as explicit beliefs that are consciously applied in social judgment and decision making. Just as importantly from a consequential standpoint, stereotypes can function as implicit, associative networks that

118. *Id.* at 28.

119. *Id.*

120. *See, e.g.,* *Anderson v. City of Bessemer City*, 470 U.S. 564 (1985) (alleging sex discrimination in employer’s failure or refusal to hire a female applicant for a traditionally male job); *Price Waterhouse v. Hopkins*, 490 U.S. 228 (1989) (finding employer liable for sex discrimination where evidence showed that some decision makers were motivated by plaintiff’s failure to conform to prescriptive sex stereotypes).

121. *See, e.g.,* *Phillips v. Martin Marietta Corp.*, 400 U.S. 542 (1971) (noting that employer may be held liable for “sex-plus” discrimination when it uses female applicants’ status as mothers of young children as a proxy for job commitment and availability).

subconsciously predispose the stereotype holder to perceive, characterize, and behave toward a stereotyped target in stereotype-consistent ways.

At a very early age, young Americans learn the stereotypes associated with the various major social groups.¹²² These stereotypes generally have a long history of repeated activation, and are apt to be highly accessible, whether or not they are believed. In other words, one can be “nonprejudiced” as a matter of conscious belief and yet remain vulnerable to the subtle cognitive and behavioral effects of implicit stereotypes. Indeed, recent neurological research using functional magnetic resonance imaging (fMRI) technology reveals that subjects who score low on measures of explicit prejudice, but whose performance on the Implicit Association Test (IAT)¹²³ reveals negative implicit attitudes toward Black targets, exhibited both a heightened startle response (measured by EMG eyeblink potentiation) and elevated amygdala¹²⁴ activation when viewing photographs of unfamiliar Black faces.¹²⁵

Left uncorrected through motivated, effortful mental processing,¹²⁶ implicit stereotypes can exercise a powerful effect on social perception and judgment. As social psychologists John Bargh and James Uleman, among others, have demonstrated, merely encountering a member of a stereotyped group primes the trait constructs associated with and, in a sense, constituting, the stereotype. Once activated, these constructs can function as implicit expectancies, spontaneously shaping the perceiver’s perception, characterization, memory, and judgment of the stereotyped target.¹²⁷ The

122. See generally Patricia G. Devine, *Automatic and Controlled Processes in Prejudice: The Role of Stereotypes and Personal Beliefs*, in *ATTITUDE STRUCTURE & FUNCTION* 181, 182-84 (Anthony R. Pratkanis et al. eds., 1989). See also MARY ELLEN GOODMAN, *RACE AWARENESS IN YOUNG CHILDREN* (rev. ed. 1964); Phyllis A. Katz, *The Acquisition of Racial Attitudes in Children*, in *TOWARD THE ELIMINATION OF RACISM* 125 (Phyllis A. Katz ed., 1976).

123. For a description and discussion of the Implicit Association Test, see generally Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 *J. PERSONALITY & SOC. PSYCHOL.* 1464 (1998).

124. The amygdala is a small bilateral structure, each half about the size of an almond, located in the temporal lobe. It plays a significant role in emotional learning and evaluation, and particularly in the evocation of emotional responses to fear or aggression-inducing stimuli. Ralph Adolphs et al., *The Human Amygdala in Social Judgment*, 393 *NATURE* 470 (1998).

125. Elizabeth A. Phelps et al., *Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation*, 12 *J. COGNITIVE NEUROSCIENCE* 729, 730.

126. An example of such motivated mental processing is setting goals to individuate the target. These goals moderate the seemingly inevitable amygdala reaction to outgroups’ faces. Mary E. Wheeler & Susan T. Fiske, *Controlling Racial Prejudice: Social-Cognitive Goals Affect Amygdala and Stereotype Activation*, 16 *PSYCHOL. SCI.* 56 (2005).

127. See generally Bargh, *supra* note 95; Mark Chen & John A. Bargh, *Nonconscious Behavioral Confirmation Processes: The Self-Fulfilling Consequences of Automatic Stereotype Activation*, 33 *J. EXPERIMENTAL SOC. PSYCHOL.* 541 (1997); James S. Uleman et al., *People as Flexible Interpreters: Evidence and Issues From Spontaneous Trait Inference*, in 28 *ADVANCES IN EXPERIMENTAL SOC. PSYCHOL.* 211 (Mark P. Zanna ed., 1996) (describing and reviewing literature on spontaneous trait inference); John A. Bargh, *The Four Horsemen of Automaticity: Awareness, Intention, Efficiency, and Control in Social Cognition*, in *HANDBOOK OF SOCIAL COGNITION* 1, 1-2

science of implicit bias demonstrates that disparate treatment can result not only from the deliberate application of consciously endorsed prejudiced beliefs, but also from the unwitting and uncorrected influence of implicit attitudes and associations in the social-perception process.

The enormous body of research examining the influence of implicit stereotypes on social judgment yields a set of key empirical findings that challenge the conception of discrimination embedded in disparate treatment doctrine. Subtle forms of intergroup bias can infiltrate decision making long before any decision is made. These biases can latently distort the perceptual data set on which that decision is ultimately premised. Often operating outside of the decision maker's attentional focus, and therefore outside his or her awareness, stereotypes can covertly but powerfully influence the way information about the stereotyped target is processed and used. They can shape the interpretation of incoming information, influence the manner in which that information is encoded into and stored in memory, and mediate the ease or difficulty with which the information is retrieved from memory and used in social judgment. A decision maker can act because of or on the basis of a target person's race, sex, or other group status, while subjectively believing that he or she is acting on the basis of some legitimate, nondiscriminatory reason.

As we will argue in Part III, Title VII's operative text prohibits these subtle forms of discrimination, but the science of implicit stereotyping has barely begun to influence federal disparate treatment jurisprudence. Indeed, from a behavioral realist standpoint, in many circuits, judicial conceptions of intergroup bias have actually regressed over the past two decades, even as psychological science has surged toward an increasingly refined understanding of the ways in which implicit prejudices bias the social judgments and choices of even well-meaning people.

One stark example of this arguably widening disconnect between the legal design of discrimination and the science of everyday prejudice can be found in a now decade-old doctrine known as the honest belief rule, which posits that to rebut an employer's nondiscriminatory explanation of a challenged employment decision, a discrimination plaintiff must prove not only that the employer's proffered reason is unworthy of credence, but also that the employer *did not honestly believe* the reason it gave.¹²⁸

(Robert S. Wyer, Jr. & Thomas K. Srull eds., 2d ed. 1994); Daniel T. Gilbert, *Thinking Lightly About Others: Automatic Components of the Social Inference Process*, in UNINTENDED THOUGHT 189, 194 (James S. Uleman & John A. Bargh eds., 1989). For a more recent treatment, discussing both automatic and controlled processes from a dual-process model perspective, see generally Timothy D. Wilson et al., *A Model of Dual Attitudes*, 107 PSYCHOL. REV. 101 (2000).

128. For a review of the honest belief-rule cases, see generally Rebecca Michaels, Note, *Legitimate Reasons for Firing: Must They Honestly Be Reasonable?*, 71 FORDHAM L. REV. 2643 (2003).

The honest belief rule, which has never been endorsed by the Supreme Court, emerged first in the Seventh Circuit¹²⁹ and later spread to the First,¹³⁰ Fourth,¹³¹ Fifth,¹³² Sixth,¹³³ Eighth,¹³⁴ Ninth,¹³⁵ Tenth,¹³⁶ Eleventh,¹³⁷ and D.C. Circuits.¹³⁸ In an early decision typical of the genre, the Seventh Circuit in *Jackson v. E.J. Brach Corp.*¹³⁹ held that a plaintiff could not prevail if the decision maker “honestly believed in the non-discriminatory reasons it offered, even if the reasons are foolish or trivial or even baseless.”¹⁴⁰ An even starker formulation can be found in *Crim v. Board of Education*,¹⁴¹ in which the court concluded that a plaintiff claiming pretext “has the burden of showing that the employer’s reason for the dismissal was a lie or had no basis in fact.”¹⁴² It is not enough, noted the court, to prove that the employer’s stated reason was doubtful or mistaken. Even if the employer’s proffered reasons were “mistaken, ill-considered or foolish, if the [defendant] honestly believed in those reasons then pretext has not been proven.”¹⁴³ The Eleventh Circuit adopted a similar approach in *Jones v.*

129. *Kariotis v. Navistar Int’l Transp. Corp.*, 131 F.3d 672, 681 (7th Cir. 1997).

130. *See, e.g., Ruiz v. Posadas de San Juan Assocs.*, 124 F.3d 243, 248 (1st Cir. 1997) (identifying the critical inquiry as “whether the employer believed that its proffered reason was credible”).

131. *Hawkins v. PepsiCo, Inc.* 203 F.3d 274, 279-80 (4th Cir. 2000), *cert. denied*, 531 U.S. 875 (2000).

132. *Deines v. Tex. Dep’t of Protective & Regulatory Servs.*, 164 F.3d 277, 280-81 (5th Cir. 1999).

133. *Braithwaite v. Timken Co.*, 258 F.3d 488, 496 (6th Cir. 2001).

134. *Eurle-Wehle v. United Parcel Serv., Inc.*, 181 F.3d 898, 900 (8th Cir. 1999) (declining to find pretext despite the doubt cast by plaintiff regarding the soundness of defendant’s proffered reason because there was no evidence indicating that the reasons were “created to disguise an illegal discriminatory motive”); *see also Vaughn v. Roadway Express, Inc.*, 164 F.3d 1087, 1090-91 (8th Cir. 1998). At least one unpublished post-*Costa* decision from the District of Minnesota has rejected the rule. *Obike v. Applied EPI, Inc.*, No. Civ. 02-1653 (JRT/FLN), 2004 WL 741657, at *4-5 (D. Minn. Mar. 24, 2004).

135. *Villiarimo v. Aloha Island Air, Inc.*, 281 F.3d 1054, 1063 (9th Cir. 2002).

136. *Evans v. Dean Foods Co.*, No. 99-4148, 2000 U.S. App. LEXIS 22519 (10th Cir. Sept. 6, 2000).

137. *Cooper v. So. Co.*, 390 F.3d 695, 729-30, 732 (11th Cir. 2004).

138. *Fischbach v. D.C. Dep’t of Corrections*, 86 F.3d 1180, 1183 (D.C. Cir. 1996).

139. 176 F.3d 971 (7th Cir. 1999).

140. *Id.* at 984.

141. 147 F.3d 535 (7th Cir. 1998).

142. *Id.* at 541 (citation omitted).

143. *Id. Accord Millbrook v. IBP, Inc.*, 280 F.3d 1169, 1179 (7th Cir. 2002) (noting that “pretext” means “lie”); *Kulumani v. Blue Cross Blue Shield Ass’n*, 224 F.3d 681, 693 (7th Cir. 2000) (noting that to prove pretext, plaintiff must establish deceit); *Pitasi v. Gartner Group, Inc.*, 184 F.3d 709, 717-18 (7th Cir. 1999) (noting that to establish pretext, plaintiff must show that the employer did not honestly believe the reasons it gave for its action); *Flores v. Preferred Technical Group*, 182 F.3d 512, 516 (7th Cir. 1999) (noting that the employer need only supply an honest reason, not necessarily a reasonable one); *Hoffman v. MCA, Inc.*, 144 F.3d 1117, 1123 (7th Cir. 1998) (noting that the key question in pretext inquiry is whether defendant honestly believes its proffered reason); *Richter v. Hook-SupeRx, Inc.*, 142 F.3d 1024, 1029-30 (7th Cir. 1998) (noting that to establish pretext, plaintiff must challenge the honesty of defendant’s proffered reason); *Jasmantas v. Suburu-lsuzu Auto.*, 139 F.3d 1155, 1157 (7th Cir. 1998) (applying the honest belief rule in an ADA disparate treatment case); *Kariotis v. Navistar Int’l Transp. Co.*, 131 F.3d 672, 676-77 (7th Cir. 1997) (noting that “if the company

Gerwens:¹⁴⁴ “The law is clear that, even if a Title VII claimant did not in fact commit the violation with which he is charged, an employer successfully rebuts any prima facie case of disparate treatment by showing that it honestly believed the employee committed the violation.”¹⁴⁵

This approach is plainly inconsistent with what empirical social psychologists have learned over the past twenty years about the manner in which stereotypes, functioning not as consciously held beliefs but as implicit expectancies, can cause a decision maker to discriminate against members of a stereotyped group.¹⁴⁶ As Darley and Gross’ Hannah study so vividly illustrated, stereotypes can bias decision making *implicitly* by skewing the manner in which inherently ambiguous information about the stereotyped target is perceived, characterized, attributed, encoded in and retrieved from memory, and used in social judgment.¹⁴⁷ Thus, it is perfectly possible for a decision maker, whose biased judgment of a negatively stereotyped target caused him to discriminate against that target, to believe that his judgment and resulting decision were based entirely on legitimate nondiscriminatory reasons. Spontaneous biased construal, often occurring outside of the decision maker’s conscious attentional focus, can easily result in the differential treatment of members of a negatively stereotyped group.

In societies or groups in which antidiscrimination norms are strong, people are highly motivated to be seen by others, and to see themselves, as nonprejudiced.¹⁴⁸ Thus, when faced with making a choice between

honestly believed in [its] reasons, the plaintiff loses even if the reasons are foolish or trivial or baseless” and “the question is not whether the employer’s reasons for a decision are ‘right,’ but whether the employer’s description of its reasons is *honest*”); *Hartley v. Wis. Bell, Inc.*, 124 F.3d 887, 890 (7th Cir. 1997) (“Plaintiffs lose if the company honestly believed in the nondiscriminatory reasons it offered, even if those reasons are foolish or trivial or even baseless.”).

144. 874 F.2d 1534 (11th Cir. 1989).

145. *Id.* at 1540.

146. This approach is also inconsistent with what social psychologists describe as “motivated reasoning.” See generally Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCHOL. BULL. 480 (1990); T. Pyszczynski & J. Greenberg, *Toward an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis-Testing Model*, in 20 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 297 (Leonard Berkowitz ed., 1987).

147. Irene V. Blair, *Implicit Stereotypes and Prejudice*, in COGNITIVE SOCIAL PSYCHOLOGY: THE PRINCETON SYMPOSIUM ON THE LEGACY AND FUTURE OF SOCIAL COGNITION 359 (Gordon B. Moskowitz ed., 2001) (reviewing the relevant literature).

148. See Philip E. Tetlock et al., *The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals*, 78 J. PERSONALITY & SOC. PSYCHOL. 853, 861-63 (2000) (discussing findings that liberals were less likely to use race-tainted stereotypes, and those who did tried to reaffirm their identity as fairminded); Wilson et al., *supra* note 127, at 100-12 (reviewing studies demonstrating that people hold both implicit, prejudiced attitudes and explicit, egalitarian attitudes, and that explicit attitudes may, with sufficient motivation, override implicit prejudices); Duane T. Wegener & Richard E. Petty, *Flexible Correction Processes in Social Judgment: The Role of Naive Theories in Corrections for Perceived Bias*, 68 J. PERSONALITY & SOC. PSYCHOL. 36, 36 (1995) (discussing how subjects adjust their judgments of targets by “flexible correction processes” when they are motivated to do so).

members of different social groups, people whose preferences are implicitly shaped by group membership spontaneously search for independent decision criteria consistent with their preference, and use those criteria to justify their choices to themselves and others.¹⁴⁹ In a powerful series of experiments, MIT business-management professor Michael Norton and his colleagues demonstrated this effect in simulated hiring and higher-education admissions decisions. They showed that subjects consistently altered the qualifications they deemed most relevant to the selection of a high-level construction manager, a stereotypically male job. When the male candidate had more education and less relevant job experience, subjects—who overwhelmingly preferred the male candidate—reported that they viewed education as more important than job experience. When the male candidate had more job experience and less education than the female candidate, subjects ranked job experience as more important than education. Either way, subjects tended to rank the criteria in a way that would justify selection of the male candidate on the grounds that he was “better qualified” than the female candidate they were rejecting.¹⁵⁰ However, when subjects were forced to rank the selection criteria before seeing the candidates’ resumes, gender bias in selection largely disappeared.¹⁵¹ Other studies demonstrate a similar “elasticity effect” through which decision makers’ judgments of the importance of various decision criteria shift in ways that justify implicit *ex ante* preferences.¹⁵²

In *The Content of Our Categories*, Krieger argued that one of the central defects of disparate treatment doctrine in general, and the pretext model of proof in particular, is the manner in which it constructs a false dichotomy between “true reasons” and “phony reasons” for allegedly discriminatory employment decisions.¹⁵³ Krieger stated that, in cases adjudicated under the pretext paradigm, the purportedly “legitimate non-discriminatory reason” proffered by the employer to explain a negative action taken against a target employee had to be classified as either a “true reason” or a “phony reason,” a “mask,” “sham,” or “cover-up” for the employer’s “true” discriminatory purpose.¹⁵⁴

Even after *Desert Palace, Inc. v. Costa*,¹⁵⁵ some courts continue to rely on the honest belief rule, despite the fact that the rule developed as

149. Michael I. Norton et al., *Casualty and Social Category Bias*, 87 J. PERSONALITY & SOC. PSYCHOL. 817, 829 (2004).

150. *Id.* at 821.

151. *Id.* at 821-22.

152. Gordon Hodson et al., *Processes in Racial Discrimination: Differential Weighting of Conflicting Information*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 460 (2002); Christopher K. Hsee, *Elastic Justification: How Unjustifiable Factors Influence Judgments*, 66 ORGANIZATIONAL BEHAV. & HUMAN DECISION PROCESSES 122 (1996).

153. See generally Krieger, *The Contents of Our Categories*, *supra* note 6, at 1178-81.

154. *Id.* at 1178.

155. *Desert Palacc, Inc. v. Costa*, 539 U.S. 90 (2003).

part of the pretext model of disparate treatment proof that the Court appeared to reject when it extended Title VII's Section 703(m)'s "motivating factor" analysis to all Title VII cases. The post-*Costa* record, albeit still limited in size, is surprisingly mixed. Courts in the Seventh and Eighth Circuits have continued to apply the honest belief rule in individual disparate treatment cases,¹⁵⁶ including those filed under Title VII.¹⁵⁷ As of this writing, only one court, in the District of Minnesota, has refused to apply the doctrine on the grounds that its analytical foundations were undermined by *Costa*.¹⁵⁸

Why should it matter that the intuitive psychology underlying the honest belief rule is unsound? The answer, of course, is that it may or may not matter, depending on the purposes being served by a court's decision whether or not to adopt it. In elaborating what it means for group status to be a motivating factor within the meaning of Title VII, courts necessarily balance a variety of competing considerations. Such considerations might include a recognition of the significant stigma associated with labeling an employer as a "discriminator" or a desire to focus scarce judicial resources on the potentially most meritorious cases. But none of the honest belief rule cases cite any such rationale in support of the rule. Rather, the logic of the honest belief rule decisions points to the operation of an unstated and unexamined judicial theory about the nature of discriminatory motivation itself—that when people discriminate they know that they are doing so. The honest belief rule assumes that a reason proffered by an employer to explain its action is either, in the words of *Price Waterhouse*, an "honest answer," or a deliberate lie.

But as we have seen, empirical cognitive social psychology has demonstrated that the lay psychological theory that underlies the honest belief rule is woefully incomplete. Stereotypes can influence social decision making through processes of subjective construal that occur outside of conscious, attentional focus. A well-meaning but implicitly biased decision maker can believe that he is basing a judgment about a member of a stereotyped group on legitimate nondiscriminatory reasons when, in fact, the target's group membership "caused" the decision maker to view the target in a unjustifiably negative light. For this reason, the honest belief rule would have little place in a behaviorally realistic disparate treatment doctrine.

156. *Sembos v. Philips Components*, 376 F.3d 696, 701, 703 (7th Cir. 2004) (applying honest belief rule in post-*Costa* ADEA case and affirming district court's grant of summary judgment to the defendant); *Jennings v. Mid-Am. Energy Co.*, 282 F. Supp. 2d 954 (S.D. Iowa, 2003) (post-*Costa* decision applying the honest belief rule in a post-*Costa* Family and Medical Leave Act case and granting summary judgment to employer).

157. *Walker v. Bd. of Regents of the Univ. of Wis. Sys.*, 329 F. Supp. 2d 1018 (W.D. Wis. 2004) (applying the honest belief rule in a post-*Costa* race- and gender-discrimination case premised on 42 U.S.C. §§ 1981 and 1983 and on Title VII).

158. *Obike v. Applied EPI, Inc.*, No. Civ. 02-1653 (JRT/FLN), 2004 WL 741657, at *4-5 (D. Minn. Mar. 24, 2004).

B. Dispositionism, Situationism, and the Same Actor Doctrine

One of the most important insights emerging from social psychology in the past fifty years is the principle that situations, along with one's subjective construal of those situations and one's attempt to negotiate the conflicting pressures they impose, exert a far more powerful effect on people's behavior than the "intuitive psychologist" generally assumes.¹⁵⁹ This principle is often referred to as "situationism," and it contrasts with the "dispositionism" that often characterizes common-sense psychological theories about what makes people act the way they do.

People usually explain other people's behavior in terms of personality traits or stable dispositions: "Why didn't Sam get his brief done on time?" "He's a procrastinator, that's why." "Why did Robert promote Bill instead of Mary?" "Are you kidding, Robert would never promote a woman; he's such a sexist." Social psychologists label this tendency "dispositionism." People tend to blame or credit other people's dispositions in explaining their behavior.¹⁶⁰ Yet empirical social psychology has repeatedly demonstrated that situations, along with people's attempts to negotiate the conflicting constraints under which they perceive themselves to be operating in those situations, exert a powerful effect on behavior.

Dispositionism has profoundly influenced the way people think about discrimination. People tend to view intergroup bias as though it were a kind of stable disposition, analogous to a personality trait, which expresses itself consistently across time and situation. Indeed, following the lead of Nobel Laureate Gary Becker, economists often describe this bias as a "taste for discrimination."¹⁶¹ Under this view, if a person is going to discriminate against members of a particular group, we expect him to do so *consistently*. We view tastes as relatively stable phenomena.

Of course, there is some merit to this view. Under certain circumstances, past behavior is a reasonably good predictor of future behavior. As recent research in social-personality psychology has shown, in some circumstances, clusters of personality traits can interact with a particular situation to predict behavior pretty reliably—in *that situation*.¹⁶² And, at

159. See Ross, *supra* note 33.

160. The tendency toward over-attribution of behavior to dispositions, often referred to as the "fundamental attribution error," appears to have a cultural component. It can be seen far more dramatically in attribution in U.S. subjects than in subjects studied in other countries, including Korea, Japan, China, and India. See generally Ara Norenzayan et al., *Cultural Similarities and Differences in Social Inference: Evidence From Behavioral Predictions and Lay Theories of Behavior*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 109 (2002) (reviewing the relevant literature).

161. Gary S. Becker, THE ECONOMICS OF DISCRIMINATION 41 (1957).

162. Yuichi Shoda et al., *Intraindividual Stability in the Organization and Patterning of Behavior: Incorporating Psychological Situations into the Idiographic Analysis of Personality*, 67 J. PERSONALITY & SOC. PSYCHOL. 674 (1994); Icek Ajzen, *Nature and Operation of Attitudes*, 52 ANN. REV. PSYCHOL. 27 (2001).

least with respect to the more extreme forms of bigotry, prejudices do tend to run in “packs.”¹⁶³

That said, people, including judges, have a tendency to overestimate the role of stable traits or tastes and to underestimate the role of situational variables in shaping social perception and behavior. As social psychologists Lee Ross and Richard Nisbett have observed, unexpected behavioral inconsistencies result from the fact that people operate in a state of tension.¹⁶⁴ Some perceptions, thoughts, incentives, and concerns pull a person in one direction; others pull him in another. Under such conditions, small changes in a situation, or in a person’s subjective interpretation of that situation, can lead to surprising changes in the person’s behavior.¹⁶⁵

Modern theory and research showing that unexpected shifts in human behavior result from the operation of dynamic mental tension systems derive primarily from work by Kurt Lewin and Leon Festinger, two of the founders of modern experimental social psychology. In the early 1950s, Lewin observed that “behavior has to be derived from a totality of coexisting facts,” that “have the character of a ‘dynamic field.’”¹⁶⁶ Because behavior results from the interaction of all of the countervailing forces bearing (or not bearing) on a person at any given time, it is simply not possible to generate a set of mechanistic laws from which one can reliably predict what particular response will result from a given stimulus.

In particular, as Festinger noted, a person’s attitudes often conflict, both internally and in relation to the attitudes of others in her social reference groups. These conflicts often place the individual in a state of tension, which can lead to quite sudden shifts in perception, beliefs, and behaviors.¹⁶⁷ As Ross and Nisbett describe, a person’s behavioral tendencies sit in a kind of “quasi-stationary equilibrium.” This equilibrium can be upset, causing unexpected behavioral results in any of the following ways: through adding or increasing behavior-impelling forces; by removing or weakening behavior-restraining ones; or some combination of the two.¹⁶⁸ Thus, while old-fashioned dominative racism might function like a stable trait or taste, there is significant reason to believe that the expression

163. See generally BOB ALTEMEYER, ENEMIES OF FREEDOM: UNDERSTANDING RIGHT-WING AUTHORITARIANISM (1988) (discussing how related attitudes and prejudices are clustered in the “authoritarian” personality); John Duckitt, *A Dual-Process Cognitive-Motivational Theory of Ideology and Prejudice*, in 33 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 41 (Mark P. Zanna ed., 2001); JIM SIDANIUS & FELICIA PRATTO, SOCIAL DOMINANCE: AN INTERGROUP THEORY OF SOCIAL HIERARCHY AND OPPRESSION (1999); Clark Freshman, *Whatever Happened to Anti-Semitism? How Social Science Theories Identify Discrimination and Promote Coalitions Between “Different” Minorities*, 85 CORNELL L. REV. 313 (2000) (reviewing research on co-occurrence of different types of discrimination in biased organizations and individuals).

164. ROSS & NISBETT, *supra* note 96, at 13.

165. *Id.* at 14-15.

166. KURT LEWIN, FIELD THEORY IN SOCIAL SCIENCE 25 (Dorwin Cartwright ed., 1951).

167. Leon Festinger, *A Theory of Social Comparison Processes*, 7 HUM. REL. 117 (1954).

168. ROSS & NISBETT, *supra* note 96.

of implicit bias is far less consistent across situations. Why would this be so?

As noted earlier, people learn at a very early age the stereotypes associated with different social groups in their society.¹⁶⁹ As a person gets older and is exposed to more egalitarian ideals, however, she may consciously reject the proposition expressed by a stereotype she has learned. However, when this occurs, neither the associations the stereotype comprises nor the emotional states that accompany them are necessarily replaced by a new enlightened attitude. Rather, the cognitive and emotional associations that formed the earlier biased attitude often continue to exist alongside the stereotype-holder's new, more enlightened conscious belief.¹⁷⁰ This leaves the social perceiver with "dual attitudes" toward members of the stereotyped group, one implicit and the other explicit.¹⁷¹

According to social psychologist Timothy Wilson and his associates, "[w]hen dual attitudes exist, the implicit attitude is activated automatically, whereas the explicit one requires more capacity and motivation to retrieve from memory."¹⁷² A person who holds a dual attitude will react to a stereotyped target in dramatically different ways, depending on whether implicit or explicit attitudes dominate the perceiver's judgmental process while he or she is forming an impression of the stereotyped target. One situation may make the explicit, nonbiased belief salient, thus conducing to nondiscriminatory action conforming to or reflecting that belief. Yet another situation might potentiate implicit, biased attitudes and associations, such as those measured by the implicit association test. Once activated, the implicit attitude may bias perception and judgment of the other person. Under Wilson's dual-process model, biased implicit attitudes will have a relatively greater effect on an actor's behavior either when there is little opportunity for controlled processing, or when in a particular situational context, the actor does not view his behavior as expressing his attitude toward a stereotyped group.¹⁷³

169. See discussion *infra* p. 55 and note 122.

170. See, e.g., Wilson et al., *supra* note 127; Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 6-7 (1989).

171. Wilson et al., *supra* note 127, at 104.

172. *Id.*

173. Timothy Wilson's dual-process approach to implicit and explicit attitudes is similar to the "MODE (Motive and Opportunity as DEterminants) Model" developed by Russell Fazio. According to this model, research indicating that people often express quite different attitudes toward the same attitude object in differing contexts can be explained by reference to the differing circumstances under which implicit versus explicit attitudes dominate social information processing. Explicit attitudes tend to influence behavior when people have both the motivation and the opportunity to engage in conscious, deliberative thought. On the other hand, when behavioral responses are spontaneous, or when people lack the opportunity or motivation to deliberate, implicit attitudes play a more dominant role. Russell H. Fazio, *Multiple Processes by Which Attitudes Guide Behavior: The MODE Model as an Integrative Framework*, in 23 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 75 (Mark. P. Zanna ed., 1990).

Although one might expect that explicit and implicit attitudes toward a particular subject would be strongly positively correlated, this is not always the case. While the degree of positive covariation is relatively high with respect to attitudes toward objects that are not socially sensitive (like dentists or snakes), this is not true with respect to dual attitudes about socially sensitive subjects, such as Whites' attitudes toward Blacks.¹⁷⁴ For example, in a meta-analysis of twenty-seven experiments involving a total of 1,562 White subjects, social psychologist John Dovidio found only a modest relationship between explicit measures of racial attitudes toward Blacks and implicit attitudes as measured by the IAT, other response latency tests, facial EMG, and galvanic skin response techniques.¹⁷⁵

As Greenwald and Krieger describe in the science overview appearing earlier in this Symposium,¹⁷⁶ implicit bias toward members of devalued groups often contradicts people's conscious egalitarian attitudes, beliefs, and self-concepts.¹⁷⁷ A self-consciously egalitarian decision maker who harbors implicit bias against members of a stereotyped or otherwise devalued group necessarily relates to members of that group through a tension system. Changes, even small changes, in the situation in which the implicitly biased but self-consciously egalitarian decision maker perceives, evaluates, and makes decisions about a stereotyped group member can profoundly alter the decision maker's judgment and behavior.

These principles, now widely accepted in the psychological sciences, have been almost completely overlooked in the judicial elaboration of Title VII's disparate treatment doctrine. As we have already observed in our discussion of the honest belief rule, disparate treatment doctrine expects biased decision makers to discriminate *consistently*.

Another judicial manifestation of this expectation can be found in a doctrine known as the "same actor inference," which emerged in 1991 in the Fourth Circuit's decision in *Proud v. Stone*,¹⁷⁸ an age discrimination case. Mr. Proud, a sixty-eight-year-old accountant, was hired in the spring of 1985 to serve as a Chief Accountant in the Army's Central Accounting Division in Ludwigsburg, Germany. Robert Klauss, who selected Proud for the position, had considered other applicants for the position, but ultimately selected Proud due to his superior background, education, and experience.

174. John F. Dovidio & Russell H. Fazio, *New Technologies for the Direct and Indirect Assessment of Attitudes*, in QUESTIONS ABOUT QUESTIONS: INQUIRIES INTO THE COGNITIVE BASES OF SURVEYS 204, 220-21 (Judith M. Tanur ed., 1992); John F. Dovidio et al., *Implicit and Explicit Attitudes: Examination of the Relationship between Measures of Intergroup Bias*, in BLACKWELL HANDBOOK OF SOCIAL PSYCHOLOGY: INTERGROUP PROCESSES 175, 177, 183 (Rupert Brown & Samuel L. Gaertner eds., 2001).

175. Dovidio et al., *supra* note 174, at 183-86.

176. Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations* 94 CALIF. L. REV. 945 (2006).

177. *Id.* at X.

178. 945 F.2d 796 (4th Cir. 1991).

In considering the applicants, the court noted, Mr. Klauss listed them on a chart, which included a column noting each applicant's age.

Proud began his job in mid-June. He never performed the particular duties for which he was hired, however, because immediately after he started working, one of his colleagues resigned, and he agreed to assume her responsibilities on a temporary basis. In short order, Mr. Klauss became dissatisfied with Proud's handling of these accounts, and by mid-October 1985, Klauss recommended that Proud be fired, which he was.

At trial, Proud presented substantial evidence that his age was a motivating factor in Klauss's decision to terminate him. He showed that Klauss based the termination recommendation on conduct that was not only consistent with standard accounting practices, but also with relevant Army regulations. He showed that younger accountants had done the same thing and were not disciplined for it and that he was replaced by a younger individual. He also presented evidence indicating that the delays in his production of certain reports, mentioned by Klauss in his termination recommendation, had been caused by several difficulties in acquiring necessary data. Not only were some of the provided data erroneous, but Proud had not been trained for these particular job duties. They were not in his original job description, and Proud had merely agreed to assume them temporarily.

In affirming the district court's decision to grant the defendant judgment as a matter of law, the Fourth Circuit stated that it did not even need to consider the plaintiff's evidence of discriminatory motivation because one fact overwhelmed all others: the person who made the decision to fire Mr. Proud had, only five months earlier, made the decision to hire him. Judge Wilkinson wrote for the panel:

In assessing whether Proud established that age was a motivating factor for his discharge, we focus on the undisputed fact that the individual who fired Proud is the same individual who hired him less than six months earlier with full knowledge of his age. One is quickly drawn to the realization that "[c]laims that employer animus exists in termination but not in hiring seem irrational." From the standpoint of the putative discriminator, "[i]t hardly makes sense to hire workers from a group one dislikes (thereby incurring the psychological costs of associating with them), only to fire them once they are on the job."¹⁷⁹

Judge Wilkinson continued:

Therefore, in cases where the hirer and the firer are the same individual and the termination of employment occurs within a relatively short time span following the hiring, *a strong inference*

179. *Id.* at 797 (citing John J. Donohue III & Petcr Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 STAN. L. REV. 983, 1017 (1991)).

exists that discrimination was not a determining factor for the adverse action taken by the employer. While we can imagine egregious facts from which a discharge in this context could still be proven to have been discriminatory, it is likely that the *compelling nature of the inference* arising from facts such as these will make cases involving this situation *amenable to resolution at an early stage*.¹⁸⁰

This portion of the opinion has come to be known as establishing the same actor inference.

Several aspects of the panel's decision here are worth noting. First, and perhaps most importantly, the panel was not simply holding that the fact that the same person who made the decision to fire Mr. Proud had earlier made the decision to hire him was *probative*¹⁸¹ of whether Mr. Proud's age influenced the firing decision and was therefore admissible to disprove discriminatory motivation. Rather, the Fourth Circuit in *Proud* found this fact is *so* probative that it relieved the court of the need to examine any of the other evidence bearing on the issue of discriminatory motivation.¹⁸² The same actor doctrine, under this formulation, functions not merely as an inference, which can be drawn or not drawn as the trier of fact deems appropriate given the facts and circumstances of the case as a whole. Instead, the *Proud* court treated the same actor rule as a judicially mandated inference—a *presumption*¹⁸³—that once raised can be rebutted only through the plaintiff's introduction of clear and convincing evidence.¹⁸⁴

180. *Id.* at 797-98 (emphases added).

181. Under Rule 401 of the Federal Rules of Evidence, evidence is "relevant" if it has "any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." FED. R. EVID. 401. In other words, evidence is relevant if it has probative value for the determination of some fact at issue in the litigation. See CHARLES TILFORD MCCORMICK, MCCORMICK ON EVIDENCE § 184 (John W. Strong ed., 5th ed. 1999). To be relevant, evidence need not conclusively prove the ultimate fact at issue. It need only have a tendency to make the existence of a fact that is of consequence to the action more or less probable than it would otherwise be. 2 JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN'S FEDERAL EVIDENCE § 401.04[2][c] (Joseph M. McLaughlin ed., 2d ed. 2005) [hereinafter WEINSTEIN'S FEDERAL EVIDENCE].

182. *Proud*, 945 F.2d at 797 ("In order to decide this case, we need not engage in a point-by-point rebuttal of each of plaintiff's manifold contentions.").

183. A presumption is a procedural rule narrowing the discretion accorded the trier of fact. Under such a rule, if a predicate fact (Fact A) is accepted by the trier of fact as true, the presumed fact (Fact B) must, as a matter of law, also be taken as true, unless rebutted by contravening evidence. WEINSTEIN'S FEDERAL EVIDENCE, *supra* note 181, § 301.02[1]. Presumptions can be rebuttable or irrebuttable, but in either case, they amount to judicially mandated inferences. GRAHAM C. LILLY, AN INTRODUCTION TO THE LAW OF EVIDENCE § 3.2 (3d ed. 1996).

184. Under Rule 301 of the Federal Rules of Evidence, the raising of a presumption imposes on the party against whom it is directed the burden of going forward with evidence to rebut or meet it. However, it does not ordinarily shift either the burden or the applicable standard of proof. See FED. R. EVID. 301; Texas Dep't of Cmty. Affairs v. Burdine, 450 U.S. 248 (1981) (noting that the raising of a presumption of discrimination by plaintiff's establishment of the elements of a prima facie case in a Title VII disparate treatment case imposes on the defendant an obligation to come forward with evidence rebutting or meeting the presumption but does not shift the burden of proof); Price

This is, in fact, how the doctrine has been constructed in many circuits. For example, in *Coghlan v. American Seafoods Co.*,¹⁸⁵ the Ninth Circuit affirmed a grant of summary judgment to a defendant in a Title VII national origin discrimination case who had based its motion for summary judgment on the same actor inference. In challenging that grant of summary judgment on appeal, the plaintiff argued¹⁸⁶ that, under long-standing Supreme Court precedent,¹⁸⁷ a trial court considering an employer's motion for summary judgment or judgment as a matter of law is not permitted to draw inferences in favor of the moving party.¹⁸⁸ Because the same actor inference is, after all, an *inference*, argued the plaintiff, it had no proper place in the adjudication of a summary judgment motion.

In rejecting the plaintiff's argument, Ninth Circuit Judge O'Scannlain stated that the same actor inference was not "a mere possible conclusion for the jury to draw,"¹⁸⁹ but rather was a "'strong inference' that a court must take into account on a summary judgment motion"¹⁹⁰ and that could be overcome only by an "extraordinarily strong showing of discrimination."¹⁹¹

The Fifth Circuit took a similar approach in *Brown v. CSC Logic, Inc.*¹⁹² *CSC Logic* was an age discrimination case in which the district court had granted the defendant's motion for summary judgment. Noting that CSC's CEO Winston Kimzey had hired the plaintiff Robert Davis at age fifty-four and then fired him "only four years later,"¹⁹³ the Fifth Circuit adopted the same actor inference for the reasons expressed in *Proud*, namely that "'claims that employer animus exists in termination but not in hiring seem irrational.'"¹⁹⁴ Although, in adopting the inference, the Fifth Circuit did not rule out the possibility that a plaintiff could theoretically adduce facts sufficient to overcome the inference raised by a same actor

Waterhouse v. Hopkins, 490 U.S. 228 (1989) (noting that Plaintiff's proof in a Title VII case that sex was a motivating factor in the challenged decision shifted the burden of proving nondiscrimination to defendant; however, it did not heighten the standard of proof to one of clear and convincing evidence). As acknowledged in Federal Rules of Evidence Rule 301, Congress may provide that a higher standard of proof apply to some presumptions. However, exceptions to the preponderance of the evidence standard are rare. WEINSTEIN'S FEDERAL EVIDENCE, *supra* note 181, § 301.02[3][b].

185. 413 F.3d 1090 (9th Cir. 2005).

186. *Id.* at 1096-97.

187. *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242 (1986); *Reeves v. Sanderson Plumbing Prods., Inc.*, 530 U.S. 133 (2000).

188. *Coghlan*, 413 F.3d at 1097 (noting that in entertaining a motion for summary judgment, "the court must draw all reasonable inferences in favor of the nonmoving party," because "the drawing of legitimate inferences from the facts are jury functions, not those of a judge" (quoting *Reeves*, 530 U.S. at 150)).

189. *Coghlan*, 413 F.3d at 1098.

190. *Id.*

191. *Id.* at 1097.

192. 82 F.3d 651, 658 (5th Cir. 1996).

193. *Id.* at 658.

194. *Id.*

fact pattern, the court found that the plaintiffs in that case, and Mr. Davis in particular, had simply not succeeded in doing so.

And what was that evidence which was insufficient in the face of a same actor fact pattern to raise a genuine issue of material fact on the question of whether Kimzey was biased against plaintiff Davis because of his age? Among other evidence, plaintiffs showed that Kimzey, the “same actor” in question, had said on the occasion of Davis’s remarriage, “you don’t need to be remarrying a young woman again; you can’t even get it up.”¹⁹⁵ On various other occasions, Kimzey referred to plaintiff Davis as “an ‘old goat.’”¹⁹⁶ At a management meeting, Kimzey said to Davis, “[Y]ou just can’t remember, you’re getting too old,” and at another time, Davis testified, Kimzey asked him whether “senility was setting in.”¹⁹⁷ Despite this and other evidence of age-based stereotyping, the Fifth Circuit upheld summary judgment on the grounds that the plaintiffs failed to adequately rebut what it referred to as the “*Proud* presumption.”¹⁹⁸

Not all circuits approach same actor evidence in this way. In *Johnson v. Zema Systems Corp.*,¹⁹⁹ a race discrimination case, the Seventh Circuit characterized the same actor rule as a simple inference that a trier of fact may draw or decline to draw as it sees fit. The court based this conclusion, in part, on insights from social cognition theory, observing that “an employer might be unaware of his own stereotypical views of African-Americans at the time of hiring.”²⁰⁰ The Third Circuit takes a similar approach, holding that a same actor fact pattern “is simply evidence like any other and should not be accorded any presumptive value.”²⁰¹ Likewise, in *Wexler v. White’s Fine Furniture, Inc.*,²⁰² the Sixth Circuit held that the same actor inference is just that, an inference, which is neither judicially mandated nor appropriately drawn in favor of a party moving for summary judgment.²⁰³ Hence, different circuits continue to disagree in their assessment of the same actor doctrine.

So what can behavioral realism offer this debate? Not surprisingly, the answer is, “it depends.” Specifically, it depends on how the same actor rule is interpreted and applied, and on the functions the rule is expected to serve. Conceivably, for example, a court might design a same actor presumption to provide employers with an incentive to hire members of disfavored groups. The judicial pitch would go something like this:

195. *Id.* at 656.

196. *Id.*

197. *CSC Logic*, 82 F.3d at 656.

198. *Id.* at 658 n.25.

199. 170 F.3d 734, 745 (7th Cir. 1999).

200. *Id. Accord Kadas v. MCI Systemhouse Corp.*, 255 F.3d 359, 361 (7th Cir. 2001).

201. *Waldron v. SL Indus., Inc.*, 56 F.3d 491, 496 n.6 (3d Cir. 1995).

202. 317 F.3d 564 (6th Cir. 2003) (en banc).

203. *See id.* at 573-74.

Hire a member of a protected group (e.g., women, ethnic or racial minorities, people with disabilities, older workers), and give them a try. If they don't perform to your standards, you can fire them. And if they sue you for discrimination, the fact that you hired them in the first place will provide you with a judicially mandated presumption of nondiscrimination that can be rebutted by the plaintiff only by clear and convincing evidence, enabling you, in most cases, to prevail on summary judgment.

This may sound far-fetched, but it bears strong resemblance to one justification for the same actor rule offered by the Fourth Circuit panel in *Proud v. Stone*. Judge Wilkinson explained the benefits of the rule as follows:

Our holding advances the aims of the statute. For almost any employer, there will be cases where an individual hired for a position does not meet the employer's expectations and a termination ensues. If former employees in these situations bring ADEA claims that are allowed to proceed to trial, employers may fear that a costly suit is possible even when there are completely legitimate reasons for a discharge. When this is coupled with the fact that individuals are far more likely to bring suits for discriminatory discharge than for discriminatory failure to hire, there is a grave risk that employers who otherwise would have no bias against older workers will now refuse to hire them in order to avoid meritless but costly ADEA actions. Courts must promptly dismiss such insubstantial claims in order to prevent the statute from becoming a cure that worsens the malady of age discrimination.²⁰⁴

Under this logic, the same actor inference works as a kind of judicially created free-enterprise zone, in which civil rights enforcement would be relaxed to encourage employers to hire members of protected groups. If a rule like the same actor inference were premised on such policy grounds, the empirical validity of the doctrine's underlying premise—that people who hire members of protected groups are unlikely to be biased against them—would be analytically inconsequential. Other types of arguments would bear far more weight in any debate over the doctrine's wisdom or fairness. These might include arguments based on dignitary interests of the workers whose civil rights would be weakened, on whether such a construction of the statute satisfied established norms of statutory interpretation, or on countervailing policy concerns.

Most circuits, however, including the Fourth Circuit where the doctrine originated, justify the same actor rule on grounds of adjudicative accuracy and efficiency. An employer decision maker who has participated in a decision to hire a member of a protected group, the argument goes, is so

204. *Proud*, 945 F.2d at 798 (citation omitted).

unlikely to express bias against him at a later date that a “strong inference” or presumption of nondiscrimination is justified on forensic grounds. Why does such an inference seem so commonsensical? Because it coheres with dispositionism, a strong belief in the cross-situational consistency of expressions of bias in discriminatory behavior. However, a great deal of empirical research suggests that dispositionism, the common-sense model of behavioral consistency on which the same actor inference is based, is deeply flawed, and that human behavior is far less consistent across situations than lay people tend to believe.

Awareness within social psychology of dispositionism’s flaws emerged as early as the late 1960s, with Walter Mischel’s landmark book, *Personality and Assessment*.²⁰⁵ Reviewing the then-existing research literature, Mischel demonstrated that across situations, most behavior shows a consistency, expressed as a Pearson’s correlation coefficient, of only about 0.3. This means (squaring the correlation coefficient to calculate the explained variance) that in the studies Mischel reviewed, only about 9% of the variance in a person’s behavior was predicted by the person’s allegedly stable personality traits.²⁰⁶

The relentless cataloging of cross-situational behavioral inconsistency that followed Mischel’s opening salvo prompted two types of responses among personality researchers. One response was to aggregate observations across many instances of behavior, thus reducing the effects of context and measurement error.²⁰⁷ Using this method, some researchers reached behavior-behavior correlations as high as 0.80 (a respectable 64% of variance in behavior predicted by past behavior). However, to achieve correlations this high required averaging across many separate observations and glossing over sometimes stark situational inconsistencies.

A second response to Mischel’s findings involved abandoning the effort to find broad cross-situational behavioral consistency and substituting a search for behavioral consistency within similar types of situations. Researchers taking this approach created a set of contingent propositions (if...then) that could detect the interplay of behavior and situation.²⁰⁸ For

205. WALTER MISCHEL, *PERSONALITY AND ASSESSMENT* (1968).

206. To illustrate, one of the studies included in Mischel’s review was conducted in the 1920s by Theodore M. Newcomb. In this study of fifty-one boys in twenty-one situations in summer camp, the degree of extraversion or introversion was measured by concrete behaviors such as talking time, time spent playing alone, and the like. The average correlation among behaviors across situations was 0.14. Squaring this value reveals that less than 2% of the variance in subjects’ behavior could be explained by the purported trait of introversion/extroversion. THEODORE M. NEWCOMB, *CONSISTENCY OF CERTAIN EXTROVERT-INTROVERT BEHAVIOR PATTERNS IN 51 PROBLEM BOYS* (1929).

207. Seymour Epstein, *The Stability of Behavior: I. On Predicting Most of the People Much of the Time*, 37 *J. OF PERSONALITY & SOC. PSYCHOL.* 1097-126 (1979).

208. Walter Mischel, *Toward an Integrative Science of the Person*, 55 *ANN. REV. OF PSYCHOL.* 1 (2004); Walter Mischel & Yuichi Shoda, *A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure*, 102 *PSYCHOL. REV.* 246 (1995); Yuichi Shoda, et al., *Links Between Personality Judgments*

example, one person might behave “aggressively” only in the narrow sense of verbally bullying subordinates, whereas another might behave “aggressively” only in physical fights in a bar. While research along these lines demonstrated that behavioral consistency could be found at meaningful levels within similar situational contexts, the same could not be said of behavior across differing situations.

So far, we have discussed research indicating that people’s past behavior poorly predicts their future behavior in different situations. Along similar lines, another body of research shows that people’s attitudes (more specific than their traits or dispositions in general) also fail to predict behavior consistently across situations. In the late 1960s, A. W. Wicker reviewed the existing research correlating attitudes and attitude-relevant behavior. He found that the correspondence between verbally expressed attitudes and overt attitude-relevant behavior was only about 0.3, as measured by Pearson’s correlation.²⁰⁹ This revelation shocked attitude researchers, who had made the common-sense assumption that people’s attitudes would accurately predict their behavior.

Attitude researchers reacted to Wicker’s work in much the same ways that personality researchers reacted to Mischel’s. By aggregating behavioral responses across many trials, researchers were able to find a reasonably respectable attitudinal “signal” within the ambient situational “noise.”²¹⁰ However, detecting this signal required averaging across many observations, among which there could be significant situation-to-situation variance. In short, this research showed that attitudes could predict behavior in certain circumstances, but only if enough observations were aggregated.

Other researchers further refined their approach to understanding the complex relationships between behaviors, attitudes, and situational variables by creating a complex predictive equation that predicted future behavior as a function of expressed attitudes, salient social norms, and the actor’s perceived control over his behavior. These factors together did a far better job of predicting future behavior than any of them alone.²¹¹

In short, whether attitudes predict behavior depends on the attitude, the context, and the person.²¹² Some attitudes (e.g., strong, unambivalent ones), some contexts (e.g., those emphasizing acting on one’s values), and

and Contextualized Behavior Patterns: Situation-Behavior Profiles of Personality Prototypes, 11 Soc. COGNITION 399 (1993).

209. A. W. Wicker, *Attitudes Versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects*, 25 J. OF SOC. ISSUES 41 (1969).

210. Ajzen, *supra* note 162.

211. Icek Ajzen & Martin Fishbein, *Attitudes and the Attitude-Behavior Relation: Reasoned and Automatic Processes*, in EUR. REV. OF SOC. PSYCHOL. 11 (Wolfgang Stroebe & Miles Hewstone eds., 2000).

212. FISKE, *supra* note 74, ch. 6.

some people (those for whom values are especially important) show higher attitude-behavior correlations than others (e.g., weak attitudes in norm-driven contexts for socially-oriented people). The relevant point here is that even professional psychological scientists have to jump through many measurement hoops to detect stability in the expression of a person's underlying attitudes. Just as the law cannot take for granted that past behavior will predict future behavior in a different context, so it cannot take for granted that attitudes, even strongly held ones, will be consistently expressed across different situational contexts.

For present purposes, the important point is that bias, understood as an attitude toward members of a particular group, only predicts overt discrimination under certain circumstances. Whether biased decision makers will act in a way that expresses their bias varies as a function of many different variables, including the influence of social norms, the extent to which particular social norms are made salient in particular situations, decision makers' perceptions of control, their motivation to avoid biased decision making, and the apparently relevant information they have at their disposal.²¹³

In general, people will judge others only when they feel entitled to do so, either on the basis of their position or on the basis of having sufficient information.²¹⁴ Thus, people tend to refrain from using their stereotypes, or making any judgment at all, when they lack what they perceive to be decision-relevant information. However, possessing even the illusion of information is sufficient to free people to make judgments. For example, Belgian researcher Vincent Yzerbyt and colleagues showed that people made more extreme and confident stereotypic judgments when they were told that a flash of light represented subliminal individuating information about the decision target—even though it was indeed merely a flash of light.²¹⁵

John Darley and Paget Gross's "Hannah" study, described in Part II.A, also suggests that the intuitive psychology underlying the same actor inference lacks validity. Darley and Gross's findings vividly demonstrate that social perceivers' resistance to or unintended reliance on implicit stereotypes varies dramatically depending on whether a particular situation makes antidiscrimination norms salient. People do not believe that baldly making social predictions or judgments based on stereotypes is appropriate, and they do not want to be perceived by others as doing so. For example,

213. Susan T. Fiske & Steven L. Neuberg, *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 SOC. PSYCHOL.* 1 (Mark P. Zanna ed., 1990).

214. Jacques-Philippe Leyens, et al., *The Social Judgeability Approach to Stereotypes*, in 3 *EUR. REV. OF SOC. PSYCHOL.* 91 (Wolfgang Stroebe & Miles Hewstone eds., 1992).

215. Vincent Y. Yzerbyt, et al., *Social Judgeability: The Impact of Meta-Informational Cues on the Use of Stereotypes*, 66 *J. OF PERSONALITY AND SOC. PSYCHOL.* 48 (1994).

when subjects were asked to make a prediction based on socio-economic information alone, Darley and Gross found that the subjects tended to protest. If forced to proceed, they tended to keep their predictions close to a neutral mean. Nonetheless, the same stereotypes eschewed by subjects when they were presented transparently readily found expression in social evaluation tasks performed in an ostensibly more implicit or "information rich" context.

The implications of this study for the same actor inference are both obvious and powerful. First, the study suggests that intergroup bias does not function as a stable trait or preference across all situations in a workplace. Whether implicit bias will result in discriminatory behavior turns on characteristics of the particular situation in which a social perceiver finds herself, the way she construes that situation, and the way she understands the conflicting constraints on her behavior within the situation. Small changes in the situational context of judgment and choice can give rise to marked behavioral inconsistency.

There are well-founded reasons for believing that implicit bias will express itself less readily in the hiring context than later in the employment relationship. In particular, the hiring context tends to make equal employment opportunity (EEO) norms and goals salient. As such, managers may be more vigilant about inhibiting responses based on stereotypes or other implicit attitudes. Moreover, during the hiring process, human-resources specialists or EEO managers may play a role in selecting applicants for a "short list," may be present at interviews, or may review decisions for compliance with the employer's EEO policies and goals. Where this occurs, the person who actually makes the hiring decision may be influenced in ways that blunt the effects of any implicit stereotypes he holds. However, this influence may wane as time goes on and equal opportunity goals become less prominent.

Accurate prediction of an applicant's future performance is often difficult because the hiring decision is based on the minimal information contained in a resume or gleaned from a brief interview. In these circumstances, employment decision makers, like Darley and Gross's subjects, may forbear from making decisions based on stereotypes. However, once the selected employee begins performing, the decision maker, now a performance evaluator, may be influenced by implicit stereotypes in the same ways Darley and Gross's subjects were. In the context of the information rich day-to-day evaluation of often ambiguous employee behaviors, equal opportunity goals are generally not salient, and decision makers are likely to be operating under conditions of relatively greater cognitive busyness. They can often believe they are basing tepid or outright negative evaluations of stereotyped employees exclusively on "legitimate

non-discriminatory reasons,²¹⁶ rather than on perceptions and judgments biased by implicit stereotypes.

There is, in short, little reason to believe that an implicitly biased employment decision maker who has hired a stereotyped person will necessarily succeed in keeping his or her subsequent evaluations of that person's performance free from the influence of implicit stereotypes. Implicit stereotypes are latent infiltrators of social perception and judgment. Decision makers who hold egalitarian beliefs but are affected by implicit bias operate in the very type of tension system that Ross and Nisbett describe. Subtle differences in the social, physical, or organizational environment can be expected to introduce a great deal of inconsistency into such a decision makers' perceptions, judgments, and decisions about the stereotyped target. Under these conditions, cross-situational consistency in intergroup perception and judgment simply cannot be expected to rise to the level required to support a presumption, or even a "strong inference," like the same actor rule. As is the case with so much of federal disparate treatment jurisprudence, the same actor inference is based on an intuitive psychological model that has been disconfirmed by advances in cognitive social psychology.

III

A BEHAVIORAL REALIST RECONSTRUCTION OF INDIVIDUAL DISPARATE TREATMENT THEORY: DEFINING AND (DIS)PROVING DISCRIMINATORY MOTIVATION

What exactly would it mean to use insights from the science of implicit bias to reconstruct Title VII disparate treatment theory from a behavioral realist perspective? Would it require the overruling of Supreme Court precedents, or, absent that, a statutory amendment? Would it, as two commentators have recently suggested, amount to a judicially sanctioned presumption that "everyone is biased," such that "[t]he problem of demonstrating actual discrimination goes away and claims of discrimination become irrefutable?"²¹⁷ And perhaps most importantly from a jurisprudential perspective, given the plain text of Title VII's statutory language, could judges interpret its operative provisions through a behavioral realist lens without exceeding their proper roles as "interpreters" as opposed to "makers" of law?

In this section, we pursue two goals. First, we seek to demonstrate that a behavioral realist interpretation of Title VII's anti-disparate treatment

216. *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973) (establishing that, in defending against a prima facie case of disparate discrimination, a defendant must articulate a legitimate non-discriminatory reason for the challenged decision).

217. Amy Wax & Philip E. Tetlock, *We're All Racists at Heart*, WALL ST. J., Dec. 1, 2005, at A16.

principle is fully consistent with the statute's plain text, with Supreme Court precedent, and with accepted, even conservative, methods of statutory interpretation. Second, we will show that a behavioral realist reading of Title VII's operative provisions would change relatively little about how disparate treatment discrimination claims are proven or defended, either in terms of the burden of proof borne by the plaintiff or in the types of evidence a plaintiff would need to adduce to prove a claim. Far from representing a radical innovation in civil rights law, a behavioral realist interpretation of the anti-disparate treatment principle would actually serve to correct the truly radical innovations represented by doctrines like the same actor inference and the honest belief rule, which are based neither in the statute's text nor on Supreme Court precedent, and which have unjustifiably narrowed Title VII's remedial reach.

A. *A Textualist Reading of Section 703(a)(1)*

As the Supreme Court stated in *Desert Palace, Inc. v. Costa*, when interpreting the meaning of Title VII, "the starting point for our analysis is the statutory text."²¹⁸ So, that is where we begin.

Title VII's Section 703(a)(1) provides in pertinent part: "It shall be an unlawful employment practice for an employer . . . to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual . . . , because of such individual's race, color, religion, sex, or national origin[.]"²¹⁹ Thus, as noted in Part I, as originally enacted in 1964, Title VII made it unlawful to "discriminate" on the basis of certain protected characteristics, but it nowhere defined precisely what the term "discrimination" meant. Moreover, it should be noted, neither the word "intentional" nor the phrase "intent to discriminate" appear anywhere in the statutory text. The intent requirement itself is a judicial innovation.

The Second Edition of the *Webster's New Collegiate Dictionary*, widely used in the early 1960s, provides no justification for the conclusion that the word "discrimination" necessarily implies conscious intentionality. That dictionary defines the verb "discriminate" as "to make a difference in treatment or favor (of one as compared with others)."²²⁰ That definition has endured: the *Webster's Third New International Dictionary of the English Language, Unabridged* defines "discrimination" similarly, if a bit more technically, as "the process by which two stimuli differing in some aspect are responded to differently."²²¹

218. *Desert Palace, Inc. v. Costa*, 539 U.S. 90, 98 (2003).

219. 78 Stat. 255, 42 U.S.C. § 2000e-2(a)(1).

220. WEBSTER'S NEW COLLEGIATE DICTIONARY 237 (2d ed. 1956) (based on WEBSTER'S NEW INTERNATIONAL DICTIONARY).

221. WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY OF THE ENGLISH LANGUAGE, UNABRIDGED 648 (2002).

There is nothing in either the text of Title VII nor the dictionary definition of the verb "discriminate" that limits the statutory text to differences in treatment resulting from an employer's conscious intention to subordinate (or favor) an individual because of his or her protected-group membership. Moreover, the Supreme Court has repeatedly acknowledged that the statute's remedial reach is broader than this crabbed interpretation. For example, the Court has repeatedly stated that, in enacting Title VII, "Congress intended to prohibit all practices in whatever form which create inequality in employment opportunity due to discrimination on the basis of race, religion, sex, or national origin."²²² On other occasions, the Court has observed, "Title VII tolerates no racial discrimination, subtle or otherwise."²²³

Given conditions prevailing in the early 1960s, it is of course reasonable to assume that conscious, purposeful disparate treatment was the most salient evil against which Title VII was directed, at least as far as race discrimination was concerned. But there is nothing in either the text of the statute or the legislative history accompanying it that would foreclose other meanings, as the Court's above-cited observations reflect.²²⁴ Thus, the notion that Section 703(a)(1)'s anti-disparate treatment principle proscribes only the deliberate, willful use of an employee or applicant's protected-group status in an employment decision is a judicial innovation, not a legislative one. What the plain language of Section 703 requires is that an employer not allow an applicant or employee's group status to play a causal role in any employment decision.²²⁵

One could nonetheless argue that in 1964, in using the term "discriminate" in Section 703(a)(1), Congress contemplated only willful bigotry, as this is arguably the way the term was most commonly understood in the early 1960s. This reading would arguably comport with the textualist principle that, in interpreting an ambiguous statutory term, one should attribute to legislators the intention "to say what one would be normally understood

222. *Alexander v. Gardner-Denver Co.*, 415 U.S. 36, 44 (1974); *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 800 (1973); *Franks v. Bowman Transp. Co.*, 424 U.S. 747, 763 (1976); *see also* *Griggs v. Duke Power Co.*, 401 U.S. 424, 429-30 (1971) (applying this principle in establishing the disparate impact theory of Title VII liability).

223. *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 526 (1993) (quoting *Green*, 411 U.S. at 801); *Price Waterhouse v. Hopkins*, 490 U.S. 228, 243 (1989) (same); *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 349 n.31 (1977) (same); *McDonald v. Santa Fe Trail Transp. Co.*, 427 U.S. 273, 280 n.8 (1976) (same).

224. As Justice Scalia observed in *Oncale v. Sundowner Offshore Servs.*, 523 U.S. 75, 79 (1998), "statutory prohibitions often go beyond the principal evil to cover reasonably comparable evils, and it is ultimately the provisions of our laws rather than the principal concerns of our legislators by which we are governed."

225. Professor Krieger, writing with Professor Rebecca Hanner White, has advanced this argument in detail elsewhere, and so, for purposes of brevity, it will not be further developed here. White & Krieger, *supra* note 21.

as saying, given the circumstances in which one said it.”²²⁶ Thus, the argument would go, in 1964, the term “discrimination” generally was understood to mean “intentional discrimination,” with “intent” defined as synonymous with “design” or “purpose.”²²⁷ Such an interpretation, were it accepted, would arguably exclude from the statute’s reach disparate treatment based on implicit stereotypes, or other implicit mental processes or associations of which the decision maker was not aware and did not deliberately set out to manifest.

This argument would confront a number of serious problems, however. The first lies in the semantic structure of Section 703(a)(1) itself. As described earlier, the section provides, among other things, that it is an unlawful employment practice for an employer to “fail or refuse”²²⁸ to hire an individual because of his or her protected-group status. “Fail” and “refuse” have very different meanings. In the early 1960s “to fail” was commonly defined as meaning “to fall short,” “to err; to be mistaken.”²²⁹ More recent and complete dictionaries define “fail” in pertinent part as “to be deficient or unable to meet a test or standard of attainment,” or “to err in judgment: be in error,”²³⁰ “to be deficient or inadequate.”²³¹ To “refuse,” on the other hand, means “to withhold . . . compliance,” “to show or express a positive unwillingness to do or comply with.”²³² Refusal connotes willfulness, as opposed to the mere missing of the mark connoted by failure. Thus, in its very semantic structure, Section 703(a)(1) indicates that the concept of discrimination comprises two variants: one willful and deliberate, the other a mere falling short of some egalitarian goal or aspiration.

B. *Discriminatory Motivation, Congressional Intent, and the 1991 Amendment of Section 703*

The argument that Section 703(a)(1)’s use of the term “to discriminate” refers only to intentional action in its narrowest sense suffers from a second and perhaps even more damning flaw. Specifically, in a portion of the Civil Rights Act of 1991 codified as Title VII’s Section 703(m), Congress amended Section 703’s operative language, providing that “an

226. Joseph Raz, *Intention in Interpretation*, in *THE AUTONOMY OF LAW: ESSAYS ON LEGAL POSITIVISM* 249, 268 (Robert P. George ed., 1996).

227. Such a reading would be supported by the Webster’s New Collegiate Dictionary published in 1956. It defines “intent” as “purpose” or “design,” and lists as additional synonyms “aim,” “end,” “objective,” and “goal.” WEBSTER’S NEW COLLEGIATE DICTIONARY, *supra* note 220, at 438.

228. 42 U.S.C. § 2000e-2(a)(1) (2001).

229. WEBSTER’S NEW COLLEGIATE DICTIONARY, *supra* note 220, at 296.

230. WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY, *supra* note 221, at 815.

231. MERRIAM-WEBSTER, WEBSTER’S ONLINE THIRD NEW INTERNATIONAL DICTIONARY, UNABRIDGED (2002), <http://unabridged.merriam-webster.com>.

232. WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY, *supra* note 221, at 1910; MERRIAM-WEBSTER, WEBSTER’S ONLINE THIRD NEW INTERNATIONAL DICTIONARY, UNABRIDGED (2002), <http://unabridged.merriam-webster.com>.

unlawful employment practice is established when the complaining party demonstrates that race, color, religion, sex, or national origin was a motivating factor for any employment practice, even though other factors also motivated the practice."²³³

The question thus becomes, what did Congress mean when it used the term "motivating factor"? What does it mean for a protected-group characteristic, like race or sex, to "motivate" an employment decision maker, and what, if anything, is the difference between "motive" and "intent"?

Although they are similar, "motive" does not mean the same thing as "intent." *Black's Law Dictionary*, for example, distinguishes the two concepts as follows: "Motive is what prompts a person to act, or fail to act. Intent refers only to the state of mind with which the act is done or omitted."²³⁴ A distinction between intent and motive, and the implications of that distinction for the legal relevance of implicit stereotyping, has long been made in the disparate treatment jurisprudence developed under the Age Discrimination in Employment Act.²³⁵ In these cases, courts have acknowledged that an employment decision maker can, through the operation of implicit stereotypes, be "motivated" by an applicant's or employee's age, even absent a specific "intent" to discriminate on the decision maker's part.²³⁶

The *Webster's Third New International Dictionary of the English Language, Unabridged* defines the word "motive" as "something within a person (as need, idea, organic state, or emotion) that incites him to action."²³⁷ In other words, a "motivating factor" is an internal mental state, a category that includes cognitive structures like implicit stereotypes or other social schema that influence social perception, judgment, and action. For race, color, sex, national origin, or other protected characteristics to "motivate" an employment decision means that the characteristic served as a stimulus which, interacting with the decision maker's internal biased mental state, led the decision maker to behave toward the person differently than he otherwise would.

233. 42 U.S.C. § 2000e-2(m).

234. BLACK'S LAW DICTIONARY 727 (5th ed. 1979).

235. For a thorough discussion of the relevant case law, see Krieger, *The Content of Our Categories*, *supra* note 6, at 1171.

236. *Syvock v. Milwaukee Boiler Mfg. Co.*, 665 F.2d 149, 155 (7th Cir. 1981) (holding that age discrimination may arise from an unconscious application of stereotyped notions of ability); *La Montagne v. Am. Convenience Prods., Inc.*, 750 F.2d 1405, 1410 (7th Cir. 1984) ("Age discrimination may be subtle and even unconscious."); *Burlew v. Eaton Corp.*, 869 F.2d 1063, 1066 (7th Cir. 1989) (holding that age discrimination may exist absent an intent to discriminate).

237. WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY, *supra* note 221, at 1475.

C. *Behavioral Realism, Legal Normativity, and the Nature of
Discriminatory Motivation*

Given the plain meaning of the term “motivating factor,” the question then becomes, what are the various ways in which an applicant or employee’s race, sex, or other Title VII-protected characteristic might influence a covered employer’s decision-making process? This is obviously a question about human psychology, about which psychological science rightly has much to say.

Many different mental processes can cause an employment decision maker to “respond differently” to members of different social groups. At the most blatant end of the spectrum, for example, an employer might refuse to hire or promote a woman to a particular position because he believes that women should not hold jobs of the type in question. Or, the employer might refuse to hire a Latino as opposed to an Anglo because he believes that, on average, Anglos are more conscientious than Latinos, and he is using ethnicity as a proxy for future productivity. Open, deliberate application of these sorts of explicit normative and descriptive stereotypes is easily recognizable as discrimination, but it does not represent the only way in which two people differing with respect to race, gender, or some other Title VII protected trait, can be “responded to differently”²³⁸ because of that trait.

As shown in earlier work by Krieger and others,²³⁹ and as other articles in this Symposium describe,²⁴⁰ discrimination—the process by which two people differing in some respect are responded to differently—does not result only from the deliberate application of consciously endorsed prejudiced beliefs. Such differences in response can also result from the application of implicit attitudes, including implicit stereotypes—networks of mental associations that predispose the stereotype-holder to perceive, characterize, and behave toward members of different socially categorized groups in different ways.

However, just because it is *possible* that implicit stereotypes might distort perception and judgment in a way that results in discrimination does not make it *inevitable* that they will do so in any given case. As Hal Arkes and Philip Tetlock have observed,²⁴¹ and we described in detail in Part II above, context is a powerful moderator of implicit bias. Recognizing that implicit bias may cause a decision maker to perceive, judge, or behave differently toward a member of a Title VII-protected group does not relieve a

238. See WEBSTER’S NEW COLLEGIATE DICTIONARY, *supra* note 222, at 237.

239. See sources cited *supra* note 21.

240. Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of “Affirmative Action,”* 94 CALIF. L. REV. 1063 (2006); Christine Jolls & Cass Sunstein, *The Law of Implicit Bias*, 94 CALIF. L. REV. 969 (2006).

241. Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or “Would Jesse Jackson ‘Fail’ the Implicit Association Test?”* 15 PSYCHOL. INQUIRY 257, 275 (2004).

discrimination plaintiff of proving by a preponderance of the evidence that this in fact occurred in his or her case. As the Supreme Court stated in *International Brotherhood of Teamsters v. United States*,²⁴² in a Title VII disparate treatment case, “[p]roof of discriminatory motive is critical, although it can in some situations be inferred from the mere fact of differences in treatment.”²⁴³

The critical question is this: if a Title VII plaintiff is able to prove that the unwitting application of derogating race, gender, or ethnic stereotypes motivated a decision maker and thus deprived her of an employment opportunity, has she proven that her race, gender, or ethnicity was a “motivating factor” within the meaning of Title VII? Doctrines like the same actor inference and the honest belief rule, and the lay psychological theories judges use to justify them, function only to obscure rather than resolve the important normative choice this question represents.

Given the plain meaning of the operative words of Section 703, and the Supreme Court’s prior statements about the normative goals underlying Title VII,²⁴⁴ it would be difficult to justify an interpretation of the statute that excluded from its remedial reach employment discrimination resulting from the thoughtless application of derogating gender, racial, or ethnic stereotypes. Nothing in the Supreme Court’s prior precedents, particularly those postdating the amendment of Section 703 in 1991, could create a reasonable reliance interest in such a narrow reading of the anti-disparate treatment principle.

Behavioral realism, then, does not substitute empirical findings for normative legal analysis. Rather, it challenges courts not to use *faux* empiricism to avoid facing difficult normative choices head on. As David Faigman suggests, the use of insights from the empirical social sciences can often force a court “to confront the value choices it [is] making.”²⁴⁵ Seen in this way, behavioral realism conduces not to judicial activism, but to judicial restraint. Science restrains judges by holding them accountable for their rhetorical use of empirical claims in legal analysis.

D. *Proving and Disproving Discriminatory Motivation*

At this point, an astute reader might well be asking, “What about proof? What kind of evidentiary showing would be required to establish motivation if we were to adopt the theory of intergroup bias toward which behavioral realism points?” Although a thoroughgoing explication of proof of discriminatory motivation would require a much higher word limit than

242. 431 U.S. 324 (1977).

243. *Id.* at 335 n.15.

244. See *supra* text accompanying notes 215 and 217.

245. David L. Faigman, “Normative Constitutional Fact-Finding”: Exploring the Empirical Component of Constitutional Interpretation, 139 U. PA. L. REV. 541, 608 (1991).

the new law review article length guidelines permit, the question is apt and we feel compelled to provide at least some discussion of this issue.

Those who anticipate a radical change in the types of evidence litigants will proffer to prove or disprove discriminatory motivation under a behaviorally realistic interpretation of Section 703 will be disappointed (or perhaps relieved) by what we propose here. As Krieger suggested in 1995,²⁴⁶ disparate treatment resulting from the mindless application of uncorrected implicit stereotypes can be proven or disproven through the same types evidence long recognized as relevant on the question of intent in disparate treatment adjudication. Under the framework we propose, the evidence would remain much the same, but the inferences reasonably drawn from that evidence, and the nature of the ultimate fact the evidence would be offered to prove, would expand to accommodate the insight that disparate treatment can result from the uncorrected influence of implicit stereotypes as well as from their deliberate, fully conscious use.

In this regard, consider again the description in Part I of how a plaintiff in a disparate treatment case goes about establishing a valid, legally cognizable claim. To establish entitlement to legal relief, this plaintiff must introduce admissible evidence sufficient to establish the existence of each essential element of the disparate treatment claim, including the intent element—that is, that her protected group status was a “motivating factor” in the challenged employment action. However, as in most cases, which must be proved or disproved through circumstantial evidence, it is not the evidence placed into the record that establishes or disestablishes an essential element of the particular claim for relief on which it was offered. Rather, in most situations, the essential elements may be proved through the presentation of numerous evidentiary facts that, taken together with each other and with the reasonable inferences drawn from them, potentially establish the existence of each essential element by a preponderance of the evidence.

Over the past thirty-three years, the Supreme Court has described the types of evidentiary facts a trier of fact may consider in a disparate treatment case in deciding whether the plaintiff’s protected group status was a motivating factor in a challenged employment decision. These include:

- Comparative evidence showing whether similarly situated persons not in the plaintiff’s protected group were treated more favorably than the plaintiff or other members of the plaintiff’s group;²⁴⁷
- Statements or expressive conduct by decision makers evincing negative stereotypes or attitudes toward the plaintiff or others in his or her protected group;²⁴⁸

246. See generally Krieger, *The Contents of Our Categories*, *supra* note 6, at 1242.

247. *Reeves v. Sanderson Plumbing Prods., Inc.*, 530 U.S. 133, 151-52 (2000); *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 804 (1973).

248. *Reeves*, 530 U.S. at 151.

- The employer's willingness to tolerate harassment of the plaintiff or other members of the plaintiff's protected group;²⁴⁹
- The employer's general pattern of treatment of members of the plaintiff's group, including statistical evidence;²⁵⁰
- The specific decision maker's treatment of plaintiff and other members of the plaintiff's protected group;²⁵¹
- Whether the defendant has come forward with evidence of a legitimate, nondiscriminatory reason for its action;²⁵²
- Whether the nondiscriminatory reasons proffered by the defendant fail to rationally explain the decision it made or are otherwise unworthy of credence;²⁵³
- Whether the nature or operation of the employer's decision-making process left room for the operation of bias;²⁵⁴ and
- Whether the employer had in place and applied effective mechanisms for detecting the possible influence of bias and for preventing such biases from influencing the ultimate decision made.²⁵⁵

Under the behaviorally realistic approach to defining discriminatory motivation described earlier, these species of evidence are as probative as they have always been. What changes under a behavioral realist interpretation of Section 703(m) is the set of inferences that can reasonably be drawn from these species of evidence and what exactly "discriminatory motivation" means, as an essential element of the plaintiff's disparate treatment claim. For example, comparative evidence showing that an African-American plaintiff was treated more harshly than a similarly situated White employee in a disciplinary situation could be offered to show that implicit stereotypes had caused the decision maker to perceive the plaintiff's misconduct to have been more serious, reprehensible, or likely to recur, than the similar misconduct of the White comparator. Statements made by the decision maker that suggest that the decision maker harbored stereotype-infused attitudes or beliefs could bolster such an inference, whether such statements were directed toward the plaintiff or not. On the other hand, the decision maker's record of hiring or promoting other African-Americans might tip the balance the other way.

The point is that in any given case, particular pieces of evidence may not be legally sufficient to *compel* a finding of discriminatory motivation, but legal doctrine should not be so constrictive as to deprive the trier of fact

249. *Patterson v. McLean Credit Union*, 491 U.S. 164, 188 (1989) (noting proof of disparate treatment under 42 U.S.C. § 1981).

250. *Green*, 411 U.S. at 804-05.

251. *Reeves*, 530 U.S. at 151-52; *Patterson*, 491 U.S. at 188; *Green*, 411 U.S. at 804-05.

252. *Tex. Dep't of Cmty. Affairs v. Burdine*, 450 U.S. 248, 254 (1981).

253. *Reeves*, 530 U.S. at 147; *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 511 (1993).

254. *Price Waterhouse v. Hopkins*, 490 U.S. 228, 251 (1989).

255. *Id.*

of its proper role in determining whether, taken together, they are. The trier of fact's proper role is to consider all the evidence and combine it with inferences it deems reasonable in order to determine whether the challenged conduct violates the operative terms of the statute as interpreted by existing legal doctrine.

In determining in disparate treatment cases what inferences would be reasonably drawn from particular evidentiary facts, or in describing the meaning of discriminatory motivation or intent, judges are constantly using psychological theories. As we have shown, disparate treatment doctrine has been premised on numerous factual suppositions about the nature, causes, and characteristics of intergroup bias. These suppositions are neither indicated nor justified by sound principles of statutory interpretation, and in many respects, they are descriptively inaccurate. In the Title VII context, naïve psychological theories like those underlying the same actor and honest belief rules function primarily to obscure an ideologically premised and jurisprudentially unjustifiable constriction of the statute's remedial reach.

CONCLUSION

EMPIRICISM, BEHAVIORAL REALISM, AND THE NORMATIVITY OF LAW

Behavioral realism does not attempt to introduce social science into normative legal reasoning. Rather, it seeks to extract from normative legal reasoning the intuitive social science already there and to subject it to empirical scrutiny. Understood as a normative theory of adjudication, behavioral realism seeks to hold judges accountable for their rhetorical use of empirical propositions that have in fact been invalidated by advances in the empirical social sciences.

Although behavioral realism sounds in empiricism, its agenda is ultimately normative in the sense that it challenges judges to satisfy the rigors of empiricism when they deploy empirically testable claims to bolster their legal reasoning. If a legal doctrine like the same actor inference described in Part II rhetorically relies on a testable social science claim, then that claim should be open to scrutiny under empiricism's evaluative standards. When a court, like the Supreme Court in *Faragher* and *Ellerth*, justifies its adoption of an affirmative defense to otherwise meritorious harassment cases on the grounds that anti-harassment training programs and grievance procedures will prevent harassment from occurring, social science evidence on this question is of undeniable import in evaluating the defense's merit as a normative legal rule.

Antidiscrimination law has long incorporated and reified factual suppositions about the nature of prejudice. Discriminatory motivation is equated with conscious intentionality. Social decision makers are presumed to have unimpeded access to the true reasons behind the decisions they make. Social decision making is construed as a process independent of

social perception and judgment. Bias is presumed to function as a stable trait, or taste, that “resides” within an individual decision maker and expresses itself consistently over at least short periods of time, wholly independent of the particular situation in which the decision maker perceives, judges, and makes decisions about members of different social groups. But well-established insights from psychological science, accumulated over fifty years of peer-reviewed, replicated research, has called these suppositions into serious doubt, if not discredited them entirely.

We do not yet know whether social psychology will eventually provide a simple, elegant, predictive model of human behavior. But in our view, its value to law does not turn on its ability to do so. Empirical social psychology is interesting because it produces vivid, replicable demonstrations of how human beings tend to behave in particular contexts. Such demonstrations are important to law because they often sharply contradict widely accepted psychological intuitions that judges mistakenly *think* provide a simple, elegant, predictive model of human behavior.

One day, the United States Supreme Court, or some state supreme court, will be confronted with a disparate treatment case in which the fact finder has concluded that implicit stereotypes, operating outside of the decision maker’s conscious awareness, caused that decision maker to subject a negatively stereotyped plaintiff to some negative employment action. In that case, the trier of fact will have found that the decision maker was not aware that implicit bias had influenced his judgment, but it will also have found that implicit bias did in fact do so.

When that case comes before the Court for review, how should the Court reason its way to an outcome? Surely, the Court should consider the statutory text, the applicable legislative history, its earlier decisions, and other traditional methods of legal reasoning. But these legal methods often underspecify legal results. In the final analysis, the Court will face a normative choice. Behavioral realism stands for the principle that, in deciding which normative choice to make, the Court should, where possible, use psychological science, not *a priori* intuitive psychological theories, in describing, justifying, or predicting the consequences of its chosen legal rule.