

A National Report Card on Discrimination in America: The Role of Testing

**Michael Fix and
Margery Austin Turner,
Editors**



The Urban
Institute

2100 M Street, N.W.

Washington, D.C. 20037

Phone: 202-833-7200

Fax: 202-429-0687

E-Mail: paffairs@ui.urban.org

<http://www.urban.org>

Copyright © March 1998. The Urban Institute. All rights reserved. Except for short quotes, no part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by information storage or retrieval system, without written permission from The Urban Institute.

The nonpartisan Urban Institute publishes studies, reports, and books on timely topics worthy of public consideration. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders; they do not necessarily reflect the official policies or positions of the U.S. Department of Housing and Urban Development or organizations with which the authors collaborated on this project.

Acknowledgments

This volume and the conference upon which it was based were funded by the Department of Housing and Urban Development's (HUD) Office of Policy Development and Research. We extend special thanks to Christopher Edley (Harvard Law School) and John Goering (HUD) for helping to conceive and plan the conference, and for their thoughtful comments on earlier versions of this volume. Thanks also go to the conference participants, all of whom contributed valuable insights on the utility and design of a national report card on racial and ethnic discrimination. Any errors and omissions are, of course, the responsibility of the volume authors and editors.

Contents

List of Conference Attendees vii

Foreword by Andrew Cuomo xi

Executive Summary 1

Conference Overview

1. Measuring Racial and Ethnic Discrimination in America
Michael Fix and Margery Austin Turner 7

Housing

2. Testing for Discrimination in Housing and Related Markets
John Yinger 27

Employment

3. Adding Testing to the Nation's Portfolio of Information on
Employment Discrimination *Marc Bendick, Jr.* 47

Commercial Transactions

4. Racial Discrimination in "Everyday" Commercial Transactions:
What Do We Know, What Do We Need to Know, and
How Can We Find Out? *Peter Siegelman* 69

Business

5. Minority Business Development: Identification and Measurement
of Discriminatory Barriers *Timothy Bates* 99

Recent Trends

6. The Future of Civil Rights Testing: Current Trends and New Directions
Roderic V.O. Boggs 113

About the Editors 123

About the Contributors 125

List of Conference Attendees

Honorable Bill Apgar
*U.S. Department of Housing and
Urban Development*

Honorable Eva Plaza
*U.S. Department of Housing and
Urban Development*

Xavier de Souza Briggs
*U.S. Department of Housing and
Urban Development*

Paul Leonard
*U.S. Department of Housing and
Urban Development*

John Goering
*U.S. Department of Housing and
Urban Development*

Paul Gatons
*U.S. Department of Housing and
Urban Development*

Kevin Neary
*U.S. Department of Housing and
Urban Development*

Harry Carey
*U.S. Department of Housing and
Urban Development*

Mercedes Marquez
*U.S. Department of Housing and
Urban Development*

Peter Kaplan
*U.S. Department of Housing and
Urban Development*

Susan Forward
*U.S. Department of Housing and
Urban Development*

Maxine Cunningham
*U.S. Department of Housing and
Urban Development*

Larry Pearl
*U.S. Department of Housing and
Urban Development*

Ivy Davis
*U.S. Department of Housing and
Urban Development*



THE URBAN
INSTITUTE

Steve Redburn
Office of Management and Budget

Susan Carr
Office of Management and Budget

John Trasviña
U.S. Department of Justice

Joan Magana
U.S. Department of Justice

Fred Freiberg
U.S. Department of Justice

Alexander “Sandy” Ross
U.S. Department of Justice

Ginette Milanes
U.S. Department of Justice

Julie Fernandes
Domestic Policy Council

Sarah Reber
Council of Economic Advisors

Sanders Korenman
Council of Economic Advisors

William Spriggs
U.S. Department of Commerce

Ellen Vargyas
*Equal Employment Opportunity
Commission*

Dianna Johnston
*Equal Employment Opportunity
Commission*

Elaine Herskowitz
*Equal Employment Opportunity
Commission*

David Garrison
*U.S. Department of Health &
Human Services*

Ruby G. Moy
U.S. Commission on Civil Rights

Stephanie Moore
U.S. Commission on Civil Rights

Sandy Wilmire
Federal Trade Commission

Lin Liu
President’s Initiative on Race

Stephen M. Cross
*Office of the Comptroller of the
Currency*

Claudia Withers
*Fair Employment Council of Greater
Washington*

Kennington Wall
*Fair Employment Council of Greater
Washington*

S. Malveaux
*Washington Lawyers’
Committee for Civil Rights and
Urban Affairs*

Shana Smith
National Fair Housing Alliance

Charles Kamasaki
National Council of La Raza

John Rehlman
*Washington Lawyers’ Committee for
Civil Rights and Urban Affairs*

David Berenbaum
*Fair Housing Council of Greater
Washington*

Margaret Simms
*Joint Center for Political and
Economic Studies*

James Gibson
D.C. Agenda/Urban Institute

LeeAnn Lodder
*Legal Assistance Foundation of
Chicago*



Paul Sonn
NAACP Legal Defense Fund

Yvonne Williams
NAACP Legal Defense Fund

Hugh B. Price
National Urban League, Inc.

Cathy Ruckleshouse
National Employment Law Project

Peter Edelman
Georgetown University School of Law

Michal Kurlaender
Harvard University

Christopher Edley
Harvard University

George Galster
Wayne State University

Reynolds Farley
The Russell Sage Foundation

Catherine Samuels
Open Society Institute

Greg Ratliff
The MacArthur Foundation

Laurie Garduque
The MacArthur Foundation



Foreword

From the founding of HUD over 30 years ago to the present, there has been one major thread that has connected HUD activities: the desire to provide improved choices and opportunities to this nation's deprived and neediest populations. HUD has done this not only through its housing programs but, equally important, through the broad range of development initiatives that are a critical part of the core mission of this Department. Ever since its establishment under President Johnson, we have striven to address a wide range of community and employment development, consumer, and housing issues as they affect urban and metropolitan areas. HUD has additionally, under my leadership, focused increasingly on our ability to serve as a clearinghouse of ideas about best practices and knowledge related to urban opportunities.

It is under this broad mandate that HUD was pleased to sponsor an Urban Institute conference in March 1998, as well as this conference volume, focusing upon the use of new methodologies for creating a national report card on the state of racial discrimination in America. In this collection of essays, a group of eminent researchers and academics have presented their best assessments and recommendations on how we may advance the nation's understanding of how well our minority group neighbors, co-workers, and colleagues are being treated. The issues addressed include employment discrimination, consumer issues, business development, housing and lending, and other areas of economic life.

This collection is premised upon one indisputable fact: employment, housing, and consumer rights and opportunities are often inextricably linked and mutually dependent. One's job opportunities and prospects clearly affect one's ability to select housing, and discrimination in restaurants and other public places surely limits a family's ability to take full advantage of all the resources and benefits of life in their community. The federal government's responsibility

to attack and to eliminate all forms of discrimination must therefore be comparably interwoven and multifaceted.

I therefore applaud the Urban Institute for drawing much-needed attention to the links and interdependences in research on discrimination in multiple arenas of social and economic endeavor. This volume significantly advances our understanding of how agencies, foundations, and private sector partners can best proceed as we search for new ways to measure, understand, and combat discrimination in all its forms. To advance this cause, I have recently launched a major initiative for a nationwide audit of housing discrimination that will provide a national and community-based report card on the state of housing discrimination in our country. This report card will offer fair housing agencies, HUD, and local communities an essential tool to better understand the ways in which, knowingly and unknowingly, minorities and other protected groups are denied equal treatment at a time when most Americans believe that justice is a fundamental part of the covenant of full participation in the country's future.

I commend this collection to you and trust that we all can learn to do more for a better and more just society in the 21st century.

Andrew Cuomo
Secretary of Housing and Urban Development



THE URBAN
INSTITUTE

Executive Summary

Despite the fact that minorities have made substantial economic and social progress over the past 30 years, significant disadvantages based on race persist within the United States and serve as markers of continuing policy failures. Claims that the nation has achieved a color-blind society appear premature as a body of empirical and anecdotal evidence indicates that discrimination based on race and ethnicity lingers, and has not been eliminated by the nation's civil rights laws.

Evidence of discrimination has come from several sources, including analysis of aggregate employment, housing, and other data sets. While the regression techniques employed in these analyses have much to offer, they fail to provide the clear, direct measures and narrative power offered by paired testing. (In a paired test, two individuals are matched for all relevant characteristics other than the one that is expected to lead to discrimination. The testers apply for a job, an apartment or some other good and the outcomes and treatment they receive are closely monitored.) But despite their power, testing studies have only been sporadically mounted over the past two decades.

In March, 1998, the Urban Institute, with support from the Department of Housing and Urban Development (HUD), convened a conference that involved many of the best-known researchers working on the measurement of discrimination. The goals of the conference were to explore the feasibility and merits of creating a national report card on discrimination, assess the role that paired testing and other social science methodologies might play in its formulation, and identify the pilot research needed for the report card's full implementation.

Papers prepared for the workshop concluded that a national report card could achieve several policy goals. It could help set enforcement priorities for civil rights organizations, identify barriers to the achievement of important social objectives (such as promoting work among former welfare recipients), and reveal some of the interdependencies between discrimination in differing areas. One example is the effects of housing discrimination and segregation on employment outcomes.

The distinguished authors of the papers presented in this volume also concluded that the state-of-the-art expertise in paired testing is sufficiently advanced to conduct national-level tests in the areas of housing sales and rentals, entry-level hiring, and, perhaps, selected areas of public accommodations (taxi service, for example), and auto sales. The conferees agreed that the national report card should produce national-level estimates, as well as statistically significant results for selected metropolitan areas. They concurred that more developmental, exploratory work would be needed before testing could be broadly applied in a number of other areas, including mortgage and commercial lending, bonding, and differential access to selected social services. At the same time, certain important areas of economic life such as being hired for higher-skilled jobs and other transactions that are inherently complex may best be probed using aggregate data and regression techniques. In general, the report card on discrimination should present paired testing results, supplemented with analyses of readily available national data sets.

The balance of this executive summary briefly states the conclusions of the five authors who made presentations at the conference and who have provided chapters for this volume. In the overview chapter that follows, Michael Fix and Margery Austin Turner draw not only from these chapters, but also from the ensuing conference discussion to outline a rationale and a strategy for using testing methodologies as the core for a national report card on racial and ethnic discrimination.

Summary of Chapter Findings

Housing Discrimination

John Yinger, of Syracuse University's Maxwell School, explores the use of testing in the sales and rental of housing and the related transactions of obtaining mortgage loans and property insurance. Yinger contends that wide-scale testing has transformed the way that racial and ethnic discrimination in housing markets is viewed: from a comparatively abstract focus on housing prices and home ownership rates to concrete stories about the unequal treatment of two equal individuals. He argues that testing research has brought a transparency and narrative power not found in earlier research.

Yinger goes on to discuss some of the clear advantages of testing, noting that it minimizes the differences in treatment caused by variables that can go unobserved by studies employing other forms of quantitative analysis, such as multiple regression. He states that testing sheds light not just on the incidence and severity of discrimination, but the circumstances in which it occurs. Further, testing makes it possible to examine the multiple, complex forms that discrimination can take by observing many types of individual/agent behavior.

Yinger highlights several limits of testing in housing as well as other fields. One notable example is the fact that testing does not provide evidence of discrimination in general, but only the discrimination that occurs within the realm defined by the sampling frame (housing units advertised in newspapers, for example). Further, some forms of discrimination may remain concealed



even from testing: for example, the processing of documents for a loan where the filing of false applications is arguably illegal.

Yinger concludes by emphasizing the need for another national testing study of discrimination in the sale and rental of housing, one that would document for the first time the shifts in discrimination levels that have occurred since the 1988 Fair Housing Act was implemented. He also calls for small-scale applications of testing in new exploratory areas, focussing on the study of discrimination in mortgage loan approvals.

Employment Discrimination

Marc Bendick, Jr., of Bendick and Egan Economic Consultants, Inc., examines past and potential uses of testing in the area of employment. Bendick argues that testing is uniquely able to bridge individuals' intuitive and research-based understandings of the prevalence of discrimination. But despite this strength, he points out that only a modest body of employment testing studies have been conducted and that those tests have been limited in their demographic and geographic ranges. For example, half of all employment testing studies have been conducted in the Washington, D.C., labor market.

Bendick concludes by proposing an annual national report card on discrimination in employment. The report card would include the results of random national hiring audits, coupled with aggregate data broken down by race, ethnicity, and gender on earnings, unemployment rates, representation within differing occupations, the acquisition of employment credentials, and the number and character of discrimination disputes.

Public Accommodations and Everyday Commercial Transactions

Peter Siegelman, an economist and lecturer at the University of Connecticut Law School, addresses the idea of extending testing methods beyond housing and employment discrimination to everyday commercial transactions. Siegelman draws a distinction between transactions where discrimination takes the form of higher prices (car buying and TV repair) and those taking the form of the denial or degradation of services (hailing a taxi, being served in a restaurant). Siegelman suggests that the initial experiments in car buying could be replicated in other cities and that tests be supplemented with analyses of tax records of purchases.

In public accommodations, Siegelman refers to a 1997 Gallup Poll finding that 45 percent of blacks believed they had been discriminated against at least once in the past 30 days: 30 percent while shopping, 21 percent while dining out. Siegelman contends that audits are a necessary, but not sufficient, technique for determining whether and how often discrimination occurs in this type of transaction. Auditing is necessary, he believes, because it is the only objective means of detecting discriminatory treatment. At the same time, he identifies several challenges to the use of tests. One is the "low incidence/high frequency problem," as the occurrence of discrimination in routine transactions may be low but the fact that people engage in them frequently means that discrimination is commonly experienced. As a result, a larger number of tests



may be needed than in the housing or employment contexts in order to generate statistically significant results. Siegelman suggests that testing in public accommodations be accompanied by surveys that are carefully designed to document the extent of perceived discrimination as well as the behaviors that minorities engage in to avoid discriminatory behavior (such as not shopping in some suburban malls).

Barriers to Minority Firm Formation and Development

Wayne State University Professor Timothy Bates explores the application of testing to an area of economic life where it has yet to be attempted: measuring discriminatory barriers to minority firm formation and development. Discrimination within this context can take a number of forms, including differential access to commercial credit, supplier credit, bonding, and markets dominated by firms or higher-income customers.

Bates concludes that testing approaches hold the most potential in probing small business' access to finance. Because a firm's creditworthiness is shaped by the multiple attributes of the business (its age, location, industry, liquidity) as well as its owner (education credentials, experience, skills) there are too many variables to permit paired testing on the part of loan applicants. Rather, Bates suggests drawing two samples of small, comparatively new, single-owner firms. Pilot studies could then be conducted of the differing results that white- and black-owned firms obtain at the pre-application stage of commercial loans: whether they are discouraged from applying or steered to government-guaranteed forms of credit. These testing results would be supplemented by econometric techniques to develop measures of differential treatment. Similar analytic techniques could also be applied to compare loan *approvals* by samples of bona fide applicants: exploring the black/white differential in loan approvals, and, where appropriate, loan amounts, interest rates, maturity, and collateral requirements.

Expanding and Potential Uses of Testing

In the volume's concluding chapter, Roderick Boggs, the Executive Director of the Washington Lawyers' Committee for Civil Rights Under the Law, reviews the ways in which enforcement and research testing were expanded and institutionalized during the 1990s as well as the policy areas where testing might be employed in the future. Potential new directions in employment include an expansion of efforts now underway within several federal agencies that are working in cooperation with private agencies, such as Chicago's Legal Assistance Foundation, in carrying out tests of race, national origin, and discrimination against noncitizens.

In the area of housing, Boggs points not only to the expansion of HUD's Fair Housing Initiatives Program, but the reliance on testing by the Housing Section of the Justice Department's Civil Rights Division. Looking to the future, Boggs notes that testing could be more fully introduced in the Section 8 voucher program, as well as the Department of Agriculture's rural home loan program. Other policy domains that Boggs sees as logical candidates for both



research and enforcement testing include access to government-supported health and hospital services, the provision of federal loans to small farmers, small business' access to financial assistance, and differential treatment within job referral and placement programs.



Chapter 1

Measuring Racial and Ethnic Discrimination in America

MICHAEL FIX
MARGERY AUSTIN TURNER

The Rationale for a National Report Card on Racial and Ethnic Discrimination

Why is a national report card needed at a time when so much is written about race and when race is purportedly declining in significance when determining economic outcomes? Conference planners and participants identified five important contributions that a report card could make to the nation's ongoing conversation on issues of race and ethnicity:

- To promote greater public understanding of the prevalence of discrimination and its contribution to inequality;
- To help guide the strategic planning of civil rights enforcement agencies to help them meet their performance goals under the Government Performance and Results Act (GPRA);
- To assess the extent to which discrimination undermines the achievement of other important social policy objectives, such as welfare reform, expanded homeownership, or reduced youth crime;
- To examine the implications of increasing diversity to help understand changing patterns of discrimination; and
- To develop a more comprehensive portrait of the prevalence and impact of discrimination.

Each of these potential contributions is addressed in turn below. In general, we believe the introduction of a national report card would systematize the largely haphazard, infrequent efforts to measure discrimination that have taken place in and outside government over the past 15 years.

Promoting Greater Public Consensus on the Extent of Discrimination and Its Contribution to Inequality. One rationale for a national report card is that to date we have little *direct* evidence regarding the extent of discrimination and its prevalence across places, areas of economic life, and population groups. As a result, there is little consensus about the extent of discrimination, whether instances of discrimination are rising or falling, and how it relates to the economic disadvantage that continues to follow racial and ethnic lines.

Clearly, despite advances made in the wake of the civil right revolution, substantial inequities persist in the economic outcomes of minorities and whites. For example:

- The hourly earnings of black men are 65 percent those of white men (Farley 1997a);
- Black men work 77 percent as many hours as white men (Farley 1997b);
- Black men will pay more than \$1000 more for the same new car as white men (Ayres and Siegelman 1995);
- Deep disparities persist in the receipt of state and local contracts for *all* minority groups (Enchautegui et al. 1996);
- U.S. schools (Orfield and Eaton 1996) and neighborhoods (Massey and Denton 1993) are becoming more not less segregated as we approach the 21st century.

While these trends may result, at least in part, from the persistence of racial and ethnic discrimination, they do not, in and of themselves, help us gauge its extent with any accuracy. The absence of understandable and compelling information about discrimination contributes to the sharp differences in the way groups interpret patterns of inequality and the obligation of government to alter them.¹ So while 60 percent of whites think conditions for blacks have improved during the past few years, only 35 percent of blacks share those views (Davis and Smith 1996).

It is not surprising that there is so little social consensus over the contribution of discrimination to social inequality. As Peter Siegelman notes, blatant Jim Crow discrimination is largely a thing of the past and the current mode of “Have-A-Nice-Day” discrimination is harder to detect, measure, and ultimately counteract. At the same time, progress toward integration paradoxically may mask an overall decline in discrimination, as Orlando Patterson argues (Patterson 1997). That is, increased interaction between members of differing racial or ethnic groups may lead to greater friction and more perceived acts of discrimination—despite the fact that the broader trend may be toward less discrimination and fewer discriminators. Thus, while have-a-nice-day discrimination may lead to premature claims that we have achieved a color blind society, conflicts associated with progress towards integration may generate exaggerated claims of victimization. Both types of distortions, along with the misguided policies that flow from them, can be corrected by more accurate and widely understandable measures of discrimination.



Guiding Civil Rights Enforcement Policy. The absence of direct, longitudinal measures of discrimination means that policymakers often do not know where discrimination is most commonly encountered and how successful anti-discrimination interventions have been. Many of the measures that have been heavily relied upon to identify problem areas and to evaluate interventions, such as court filings or enforcement actions, are imperfect guides to action. This owes in part to the fact that discrimination has historically led to low levels of legal actions and to the selectivity inherent in such measures (Miller and Sarat 1980).

It could be argued that the nation's civil rights enforcement tools are less sophisticated than those intended to combat pollution or reduce tax non-compliance. Take, for example, environmental controls, where continuous and quite expensive monitoring of the extent of ambient pollution directs the location, type, and intensity of enforcement. The monitoring of pollution levels has led to the introduction of enforcement strategies that have been designed to generate results that are efficient from the perspective of government enforcers, regulated entities, and protected populations.

Further, this tighter linking of enforcement activities to compliance results responds to the imperatives of the Government Performance and Results Act.² By systematically assembling longitudinal data on a random sample of firms, sectors, and geographic regions, and noting changes in discrimination levels, a national report card could help civil rights enforcement agencies rationalize their budgeting and targeting efforts. The report card could also help these agencies evaluate their performance³ and generate support for shifting resources to differing enforcement initiatives.⁴

It could be argued that the need for effective anti-discrimination enforcement has risen in an era in which the scope of affirmative action policies has been circumscribed in several important fields, particularly government contracting and—in Texas and California—higher education. This retreat from affirmative action should increase the pressure that policymakers feel to ensure that people are treated as equals across sectors of economic life and that anti-discrimination policies are adequately funded and strategically targeted.

Monitoring Discrimination That Defeats Other Social Goals. Other policy goals also dictate that discrimination be monitored. Welfare reform, for example, is spurring the entry of a new cohort of low-wage, low-skilled workers into the labor force. Most of the new entrants are women; many are members of racial and ethnic minorities. The success of the policies designed to promote welfare-to-work transitions is premised upon low barriers to labor force entry, including low levels of workplace discrimination. Similarly, U.S. housing policy has historically supported and encouraged the expansion of homeownership opportunities as a means toward individual wealth accumulation, neighborhood revitalization, and social cohesion. Clearly, discrimination in home sales transactions, mortgage lending, and property insurance would undermine continued gains in homeownership nationwide. Progress on other widely shared policy goals, such as assimilation of new immigrants and productive employment of young people who are at high-risk of involvement in crime and violence, also depends upon the sustained reduction of discrimination based on race and ethnicity.



Understanding the Implications of Increasing Diversity for Patterns of Discrimination. Confusion over the contribution of discrimination to inequality is not restricted to debates centered on African Americans. High sustained levels of immigration, dominated by non-European countries, have dramatically expanded and diversified the populations that are perceived as racial and ethnic minorities in the United States.⁵ By the year 2040, about 40 percent of the population will consist of racial and ethnic minorities, with blacks constituting less than one-third of the minority population (Fix and Passel 1994). The breakdown of a black/white racial paradigm complicates any easy understanding of the ways in which discrimination operates within society, who practices it, who its victims are, and the protections that government should provide.

One issue that high levels of immigration from non-European countries presents is whether new non-white immigrants and other ethnic minorities will be subjected to discrimination in housing, employment, and other domains of daily life. In fact, recent analyses of immigrant integration routinely ascribe the differentiated or “segmented” assimilation of some groups at least in part to discrimination.⁶ Moreover, housing and employment audits carried out by the Urban Institute and the Fair Employment Council of Greater Washington provide some direct evidence to support their claims, although studies conducted to date have focused on Latinos and not immigrants, *per se* (Cross et al. 1990; Yinger 1991).

At the same time, congressional concerns about illegal immigration have led to the imposition of broad new restrictions on employment and services to illegal immigrants that may be inducing increased discrimination against foreign-looking minorities (U.S. General Accounting Office 1990). Specifically, some employers report they have chosen to hire only U.S. citizens. Employers have also mistakenly and illegally required that noncitizens present a “green card” before they can be hired, despite the fact that they must accept other types of identity documents, as well. Finally, as the country’s demographic makeup shifts, racial and ethnic minority groups may not just be the victims of discrimination, but will be its perpetrators. A recent Los Angeles survey reveals that Asians and Latinos hold more negative views of African Americans than do whites (Bobo et al. 1995).

Developing a Comprehensive Portrait of Discrimination. A report card on discrimination could play a vital public education function by simultaneously examining discrimination across several key areas of economic life (such as housing, employment, public accommodations) within specific communities. This comprehensive approach could serve a public education function by painting a more complete and powerful portrait of the role that discrimination plays in daily life than studies that touch on a single area of economic activity, and might consequently help build public support for targeted anti-discrimination enforcement activities. This multi-point examination of discrimination should also help policymakers identify communities and populations where discrimination occurs across sectors. It could be useful, then, in obtaining greater cross-agency cooperation in law enforcement.



The Role of Paired Testing in the National Report Card

A clear consensus emerged from the conference that paired testing would form the core of the report card. Other methods for measuring discrimination also offer critical information and insights, but—where it can be implemented—paired testing offers special advantages. Therefore, national level audits should proceed in housing, employment and, perhaps, public accommodations and sales. At the same time, innovative, exploratory work that employs testing in other areas of economic activity should be supported, albeit on a smaller scale. The presentation of testing results in the national report card would be supplemented by analysis of other types of data, including surveys and administrative records.

Strengths of Paired Testing. Why the emphasis on testing? Mark Bendick writes, “In a world in which stories have more power than studies, testing generates studies that are stories.” There was a general consensus at the conference that neither econometric studies, nor attitudinal surveys, nor data on changing enforcement caseloads, tell the story about discrimination’s presence—or absence—with the same power as testing studies. Testing studies involve the direct observation of the unequal treatment of equals, a simple, concrete formulation that has great narrative power. The testing methodology is transparent; no knowledge of statistics is needed to understand it. Findings from paired testing research can be intuitively understood by policymakers, the media, and the general public.

Past testing studies are valuable not just because they reveal the incidence and severity of discrimination, but because they can go further and inform by the context in which it occurs. For example, housing studies reveal that discrimination may be most intense in integrated rather than segregated neighborhoods (Turner, Struyk, and Yinger 1991). Similarly, studies of employer hiring practices find that discrimination is most common in jobs that have the greatest interaction with customers (Turner, Fix, and Struyk 1991). These studies show that the unfavorable treatment of minorities may not take the form of outright exclusion from employment or housing opportunities, but steering people to less desirable alternatives: a job on a used rather than a new car lot, or a house in a less affluent neighborhood. In addition, the studies provide not only information about differing outcomes (a job offer versus no job offer) but also about dissimilar, discouraging treatment. In sum, testing studies are valuable because they can generate both hard numbers and textured analyses of discrimination.⁷

Differences between Testing for Research and Enforcement. It is important to distinguish between testing for research and testing conducted for law enforcement purposes. The principal goals of research testing are to quantify the incidence and forms of discrimination in order to promote public understanding and identify sectors in which enforcement might be targeted. Testing for research, then, must produce generalizable results regarding discrimination for a specified unit of analysis: an industry, a metropolitan area, or the nation as a whole. To achieve these generalizable results, the tests are randomized, using an accepted sampling frame such as newspaper advertisements.



Typically, research requires large numbers of tests in order to support statistically significant comparisons. To generate reliable and objective comparisons of minority and white experiences across a large number of tests, researchers usually use highly structured recording forms, with closed-ended, “check the box”-type items.

By contrast, the purpose of an enforcement test is to establish legal violations and to correct them either through settlement or litigation. Testing for enforcement is often complaint driven, and is typically targeted to a single firm or a narrowly targeted set of firms. The small number of firms tested, and the reliance on targeting, limits the generalizability of enforcement testing’s results. Enforcement testing often requires multiple tests of a single employer or agent, but generally does not involve the large numbers of tests typical of research testing. As a consequence, enforcement testing report forms tend to be much more open-ended, requiring test partners to provide greater narrative detail, rather than checking boxes. These forms are generally analyzed pair-by-pair by a knowledgeable analyst who compares the treatment of test partners across all aspects of the encounter, including subjective as well as objective information.

Although research and enforcement testing differ in significant ways, the distinctions between the two should not be overdrawn. Both are based on the same core methodology and protocols. They differ primarily in the way tests results are recorded and analyzed. Randomized testing of large numbers of market transactions need not be limited to research—it can and should also be applied in targeting for enforcement. Moreover, research and enforcement testing can be effectively conducted in tandem, yielding *both* market-wide estimates of the incidence of discrimination *and* case-specific evidence of individual violations. This strategy was effectively implemented in the Washington metropolitan area in 1990, when the Rockefeller Foundation funded the Urban Institute to conduct research testing for entry-level employment discrimination at the same time that the Washington Lawyers’ Committee for Civil Rights launched its initiative on enforcement testing. The two efforts shared a core methodology, and results from the research effort were publicly released at the same time that the first cases were filed from the enforcement effort.⁸

Applications of Testing to Date. As the papers in this volume indicate, testing for research and enforcement has been applied to an expanding range of activities, including:

- housing sales and rentals;
- entry-level hiring;
- access to taxi service;
- bargaining practices for auto sales;
- provision of pre-application quotes for mortgage loans and homeowners’ insurance; and
- access to health care.

Housing and employment are two areas where paired testing has been particularly well developed by researchers and practitioners. HUD has twice launched national paired testing studies to measure the national incidence of discrimination in housing rentals and sales transactions. The first of these studies—the Housing Market Practices Study (HMPS)—was completed in 1977.



It involved more than 3,200 paired tests of discrimination against African Americans in the rental and sales markets of 40 major metropolitan areas. The HMPS sites were randomly selected to be nationally representative of large urban areas, and samples of advertised rental and sales units were randomly selected from major newspapers in each site (Wienck et al. 1979). As a follow-up to the national HMPS sample of black-white tests, HUD conducted exploratory testing for discrimination against Hispanics in one metropolitan area (Hakken 1979).

Ten years later, HUD built upon the HMPS experience by launching a second national audit study—the Housing Discrimination Study (HDS). This study involved 3,800 paired tests for discrimination against African Americans and Hispanics. Again, both rental and sales markets were tested in a random sample of 25 major metropolitan areas. Black-white tests were conducted in 20 of these sites and Hispanic-Anglo tests were conducted in 13 sites. The HDS methodology also involved above-average sample sizes in five metropolitan areas, which supported in-depth analysis of variations in patterns of discrimination within urban areas (Yinger 1991).

Rigorous and reliable testing methods have also been developed to measure discrimination in hiring decisions for entry-level job openings. The first systematic application of paired testing to hiring, conducted in 1989, focused on discrimination against Hispanic men applying for entry-level jobs in Chicago and San Diego. In each of these sites, approximately 150 paired tests were conducted, based on random samples of job openings advertised in the major metropolitan newspapers (Cross et al. 1990). A similar study of hiring discrimination against African American men was conducted a year later in Chicago and Washington, D.C. Again, about 200 paired tests were conducted in each metro area, based on random samples of advertised job openings (Turner, Fix, and Struyk 1991). Two hundred and eighty-five paired tests of discrimination against both Hispanic and African American men were conducted in Denver at about the same time (James and Castillo 1992). Together, these testing studies (and subsequent enforcement tests conducted by local advocacy organizations) have produced an accepted and credible methodology to test for discrimination in entry-level hiring.

Pioneering efforts by both researchers and practitioners have explored the applicability of paired testing to a number of other areas: taxicab service, car sales, access to health club membership, access to property insurance and mortgage lending. The method is easier and cheaper to implement in some (taxicabs and car sales) than others (property insurance) and these differences hold implications for the design of the studies that will serve as the foundation for the national report card, as we discuss below.

Implementation Lessons That Have Emerged from Testing to Date. The research and enforcement audits that have been mounted to date have yielded important lessons about all phases of the implementation of audits. Some of these basic understandings bear on:

- **Tester recruitment.** Audits require a pool of candidates large enough to ensure close matches of capable testers. Several dozen candidates must often be interviewed for each candidate hired. Students from local universities,



identified through networks of professors, have proved to be a good source of tester candidates.

- **Tester matching and selection.** To ensure close matches in the testing process, the *hiring* of tester candidates should be contingent on finding a close match for the applicant. Testers should be matched not only on readily observable traits (height, build, general attractiveness) but also on intangible traits such as personality, warmth, gregariousness, and the like. Convening panels to evaluate the candidates and videotaping have promoted closer matching. Personality tests, however, have not proven helpful in generating closer matches.
- **Tester training.** The role and importance of training have become clear as testing studies have evolved. Training needs to accomplish at least three goals: (1) ensuring that the testers can carry out their roles convincingly; (2) bringing about a convergence in the personal styles of the partners in each tester team; and (3) ensuring that the testers understand the need for complete objectivity, eliminating any predisposition to find discrimination.⁹ Tester training typically lasts at least one week.
- **Tester compensation.** Studies conducted to date have raised concerns about the possibility that some compensation methods may create incentives for testers to either find or not find discrimination. In addition, payment approaches should not encourage testers to draw on their own personal, idiosyncratic resources to elicit particular outcomes, thereby altering the matched style of their presentations to employers or other agents. Many researchers have concluded that testers should be paid a fixed salary, so that they are not rushing or making extra efforts to complete tests. They have also concluded that testers should waive their rights to receive any damages that flow from litigation to ensure that they have no interest in eliciting discriminatory treatment.
- **Sampling design.** In order to yield generalizable results, research testing needs to draw a sample of market actors (lenders, employers, car dealerships) or a sample of market transactions (job openings, apartments for rent). Depending upon how this sample is identified, it may not fully represent *all* relevant actors or transactions, and it may not represent those most likely to be encountered by minorities. In research testing conducted to date, metropolitan newspapers have provided effective sampling frames for job and housing vacancies, and the yellow pages have been used to sample property insurance agents. However, both these sampling frames may leave out important segments of the market, where the incidence of discrimination could differ. As testing methodologies mature, strategies for sampling from a wider range of market actors and transactions should be explored.
- **Supervision of tester performance.** A full-time site coordinator is needed to ensure that testers carry out their audits in a timely manner, present themselves in the sequence dictated by the study design, call back at prescribed intervals, and that differences in results across pairs do not arise from poor matches. Supervisors also ensure that forms are filled out completely and accurately as soon as tests are completed.¹⁰ Experience to date suggests that one site coordinator cannot effectively manage more than four teams of paired testers.



- **Strategies for limiting intrusiveness.** The imperative to limit the intrusiveness of the test has been widely recognized, as has the need to ensure that tests do not penalize bona fide candidates for the tested opportunity. In employment testing for research, these objectives have been achieved by limiting any firm to a single paired test and by having testers turn down job offers immediately. In testing car sellers, auditors visited car dealers at the least busy times of the week; testers were told to return if the sales staff was busy, and the tests themselves were designed to be completed in 10 to 15 minutes.

The Limits of Testing. Experience with testing to date has also revealed a set of lessons regarding the limits of testing, especially when carried out within the context of large-scale statistical studies. Those lessons reveal that testing may not be a particularly effective technique for measuring discrimination:

- on the part of individuals making complex choices such as hiring for jobs requiring advanced skills;
- when testers run the risk of violating the law such as making false claims on credit applications;
- for activities where the incidence of discrimination is low but the number of transactions in which individuals are involved may be high (eating in restaurants);
- when the targeted industry has adapted its practices to the possibility that it might be audited as is sometimes the case in the real estate industry; and
- when discrimination occurs in activities—such as promotions or terminations—where assessments are largely based on the decisionmaker’s prior knowledge of an individual.

Take, for example, hiring for higher level jobs where the criteria for selection are complex, difficult for outsiders to gauge, and where the hiring process can be quite protracted. The challenges posed by complexity are particularly evident in efforts to test for discrimination in the availability, pricing, and terms of homeowners’ insurance. The need to simultaneously match not only testers, but the houses for which insurance is being sought, as well as neighborhoods in which they are located, complicates the effort to administer successful audits.

Another type of complexity is introduced by what Peter Siegelman refers to as race-plus discrimination. In this circumstance, black and white customers who might be treated equally in the normal course of events receive disparate treatment only when something goes awry: when a diner complains about a restaurant’s service, for example. In some instances, efforts to elicit race-plus behaviors using testing may be feasible. It does, however, require that more variables be controlled in the experiment’s conduct than straightforward efforts to simply obtain services or goods.

There are also legal barriers to testing. Outside of testing that occurs with the consent of a specific firm, full application testing of mortgage lending appears to be foreclosed to researchers because of statutory bars to filing false credit applications.¹¹ The effective use of testing also may be blocked by the adaptations that potential targets may have made to its use. One example is the new requirement that potential renters complete applications authorizing



a credit check before a landlord will show an apartment. Finally, as Siegelman notes, testing may not suit areas of economic life that involve a large number of transactions—meals eaten in restaurants, for example. Despite a low incidence of discrimination, the frequency of the activity may still mean an individual faces discriminatory behavior comparatively often. But administering the number of controlled tests required to accurately quantify the incidence of such discrimination may not be the best use of scarce resources.

When feasible, paired testing offers special advantages as a tool for measuring the incidence and forms of discrimination. Because testing directly observes unequal treatment of equally qualified people, its findings can be particularly clear and convincing for policymakers and the public. Three decades of experience with paired testing have established its feasibility and credibility across a number of important sectors, including housing transactions (both rental and sales), entry-level hiring, large-scale consumer transactions, and some day-to-day consumer services. Nevertheless, it is important to acknowledge the limits of paired testing. It can be costly, time-consuming, and logistically complex when implemented on a large scale. In particular, testing is not applicable to complex transactions in which a very large number of individual attributes are relevant to an outcome.

A Report Card on Racial and Ethnic Discrimination: Design Issues

Building upon the findings presented by the other authors in this volume, conference participants reached a general consensus about the basic design for a national report card on racial and ethnic discrimination across major economic sectors in the United States. This design encompasses national paired testing studies in housing, employment, and possibly car sales and some retail services, accompanied by smaller, developmental testing studies in mortgage lending, business development, and other areas where testing methods are not yet well established. This approach not only provides national measures of the incidence and severity of discrimination in key areas, but also contributes to the continued development of state-of-the-art testing for discrimination. The results of audits would be supplemented by time series data collected from other sources. Examples tentatively identified would include econometric studies, attitudinal surveys, and complaint and enforcement data.

The Report Card's Core: Full-Scale National Testing Studies in Housing and Employment. The accumulated experience of researchers and practitioners described above provides enough knowledge to design and implement full-scale national testing studies in housing and employment. On the basis of experience to date (documented in the previous section) paired testing studies could be designed and launched quickly—without further exploratory work—to provide a core of national measures in these two key areas.

Emerging Methods for Paired Testing: Consumer Transactions. Paired testing methods have not yet been as systematically applied to day-to-day consumer transactions. Although exploratory testing has been effectively con-



ducted by researchers and advocates in the areas of taxicab service, car sales, health club memberships, and other consumer transactions, a standardized methodology that could be readily deployed at a national scale has not yet emerged.

Nevertheless, reliable procedures could probably be designed to test nationally for discrimination in automobile sales, an important retail transaction in the lives of most Americans, one where pilot testing has revealed a significant incidence of differential treatment (Ayres 1991). Taxicab service represents another potential area for nationwide implementation. Methods have been developed and implemented in at least one city for conducting paired tests to determine whether minorities are denied service by taxicab drivers (Ridley, Bayton, and Outtz 1989). Again, the transaction is a simple one that does not require testers to be matched on many intrinsic characteristics. Building upon existing experience, a national testing study could be designed and implemented without further developmental work.

Seeding Innovation: Developing New Testing Methods for Mortgage Lending and Other Areas. Considerably more exploratory work is needed to develop methods for measuring discrimination in other important economic transactions. As discussed by Yinger (in chapter 2), major challenges have yet to be overcome to extend the testing methodology into the formal application and underwriting stages of the mortgage lending process. While HUD has funded pilot testing studies of discrimination in the provision of homeowners' insurance, they were complicated by the challenges of matching houses and neighborhoods as well as home buyers (Wissoker, Zimmermann, and Galster 1998). Neither researchers nor practitioners have experimented with the application of paired testing methods in the area of business development (e.g., obtaining commercial credit or bonding). And there are good reasons—outlined by Siegelman—to be cautious about the extent to which paired testing can effectively measure the incidence of discrimination in day-to-day economic transactions. Public services represent a largely unexplored area, where paired testing might prove to be extremely useful in measuring discrimination on the basis of race or ethnicity, but where little or no exploratory work has been conducted.

Systematic investments should be made in these areas to determine if testing for research and enforcement is practicable, and to develop and refine testing methods, so that they can ultimately be added to the national report card. More specifically, the national report card effort should include developmental work in new areas of testing with the goal of scaling up as soon as proven methodologies are available. Thus, the scope of the national report card could gradually expand over time to encompass more important economic transactions, but resources would not be squandered by trying to implement unproven methods on a national scale.

This strategy for developing and piloting new testing methods in conjunction with national implementation of well-established approaches also offers the advantage of continuously contributing to the state-of-the-art on testing for discriminatory practices. Historically, researchers and practitioners have learned from each other's innovative testing efforts, gradually extending paired testing into new areas either through experimental enforcement testing or



through pilot research efforts. An ongoing effort to develop and pilot new testing methods, in which both researchers and practitioners have a voice, could accelerate this process of innovation and learning, significantly enhancing civil rights enforcement efforts as well as increasing the state of knowledge about the incidence and severity of discrimination in different domains.

Generating Results at National and Metropolitan Levels. Like HMPS and HDS, the report card should provide statistically reliable estimates of the incidence of discrimination for the nation's urban areas as a whole. Such national estimates can be obtained from a representative sample of metropolitan areas, and would include both central city and suburban communities.¹² Some of the metro area samples should be large enough to support in-depth exploration of variations in the incidence of discrimination. Yinger argues that, although national measures are valuable, policymakers also need to know more about the circumstances in which discrimination occurs and potential causal factors that might be susceptible to policy interventions. Thus, in-depth samples might be used to explore differences in discrimination between central cities and suburbs, in racially integrated or racially changing communities, and across income groups. One possible strategy would be to select a different sub-set of metro areas for in-depth study in each report card. This would make it possible to maintain the integrity and continuity of the national sample of metro areas, while providing more detailed local analyses for a gradually expanding set of sites.

Results Updated on a Regular Cycle. Full-scale national testing for discrimination in any given sector (housing or employment, for example) probably cannot realistically be conducted more frequently than once every five years. The costs and logistical complexity of a large-scale paired-testing effort are substantial. And because the incidence of discrimination is not likely to change rapidly, measuring change in discrimination by sector on a five-year schedule should provide sufficiently current data on its incidence and severity. But not all sectors should be tested in the same year, and exploratory or pilot testing efforts for new areas should be initiated on an ongoing basis. Thus, new findings should be releasable every year (after an initial start-up period), with five-year reports that would document changes in the incidence and severity of discrimination for each sector in which testing was conducted nationally.

Supplementing Paired Testing Results with Other Data. While paired testing results should form the core of the national report card, they can and should be supplemented and extended by other types of data and analysis. For example, testing evidence on the incidence of discrimination by real estate agents in home sales might be supplemented by statistical analysis of disparate outcomes in property insurance policies or by regression analysis of differential treatment by mortgage lenders. Similarly, testing evidence of entry-level hiring discrimination might be extended and strengthened with statistical analysis of employer recruitment, hiring, and compensation practices for higher level positions. And testing evidence of discrimination in appliance repair services might be complemented by survey data reporting the frequency with which racial and ethnic minorities perceive that they have experienced discrimination in retail transactions. Hybrid approaches such as these will strengthen the overall evidence about the extent to which racial and ethnic discrimination persists across important sectors of the U.S. economy. It will also help advance the state-of-the-



art for linking testing evidence to survey data and statistical analyses of disparate outcomes, both for research and enforcement purposes.¹³

Beyond these kinds of data analysis, Bates suggests that researchers probing discrimination in commercial lending explore the use of *non-paired testing*. This would involve finding a pool of actual candidates for commercial loans. The applicants would then file genuine loan applications and the progress that they make through the loan application and approval process would be monitored and documented. The analysis would then focus on differential treatment of applicants from differing racial and ethnic backgrounds in loan approvals and, in the case of approved loans, in the loan amount, interest rates, maturity, loan type and collateral. As we have noted, legal barriers to false credit applications may rule out the use of testers not genuinely seeking loans, and would require reliance on appropriate econometric controls to assess the progress made by the bona fide loan applicants who participate in the study. The same non-paired testing approach has also been suggested for home mortgage lending (Ross and Yinger, forthcoming), and might also prove feasible in assessing discrimination in higher-skilled jobs.

Design Challenges for a National Report Card

Although the conference produced a broad consensus on the basic design of a national report card on racial and ethnic discrimination in America, significant design issues still must be resolved. These issues include developing statistical sampling procedures, building implementation capacity, balancing research and enforcement goals in testing, and determining which racial or ethnic groups to include in the report card.

Sampling Challenges. The sampling plan for the national report card needs to balance two competing objectives. It must support statistically significant generalizations about the incidence of discrimination nationwide and over time, *and* it should support in-depth analysis of discrimination across sectors in individual metro areas. Experts need to develop a stratified sampling strategy for selecting the metropolitan areas that can produce statistically valid, national measures from a manageable number of sites. In order to support rigorous analysis of changes in discrimination over time, the sample of metropolitan areas probably must remain the same for every national report card.

National testing studies conducted in the past have produced nationally reliable measures, but they were not explicitly designed to support analysis of changes in discrimination over time. Presumably, any changes in the incidence and severity of discrimination that occur at the national scale over a five-year period will be relatively small. Small changes may not be discernible as statistically significant, particularly if the variation between or within metropolitan areas is large. Therefore, the sampling plan for the national report card may require larger samples of sites and/or transactions than past studies in order to yield statistically reliable measures of *changes* in discrimination over time.

Theoretically, enough tests could be conducted in every sampled area to support statistically reliable conclusions about discrimination within individual metropolitan areas as well as nationwide. However, this would probably



be very costly. Therefore, as discussed earlier, the sampling plans for the national report card should allow for expanded samples in a subset of metropolitan areas. These expanded samples should be large enough to support statistically significant measures of discrimination at the metropolitan level, and to analyze variations in the incidence and severity of discrimination between geographic areas, household types, and types of businesses within the metropolitan region.

Using these expanded samples, analysts could report on the incidence of discrimination in particular metropolitan areas, focusing the attention of local policymakers and advocates. In addition, researchers could conduct exploratory analyses of potential variations in the incidence of discrimination and interactions between levels of discrimination in different domains. Sample design experts should explore the feasibility of targeting the expanded samples to a different subset of metropolitan areas every five years. In other words, the overall sample of metropolitan areas would remain fixed, but the areas highlighted for exploratory analyses would change. Over time, this strategy would have the advantage of focusing the attention of local policymakers in the largest possible number of metropolitan areas. It would not, however, support analysis of changes in the incidence of discrimination over time on a site-by-site basis.

In addition to the challenge of selecting metropolitan areas and determining the number of tests to conduct in each, sampling plans for each component of the national report card will have to define a reasonable “window” on the universe of transactions for which discrimination levels are to be measured. To illustrate, paired testing studies of rental and sales markets have typically sampled from houses and apartments advertised in major metropolitan newspapers. Similarly, testing studies of employment discrimination have sampled from job openings advertised in the newspaper. A recent exploratory study of discrimination in the provision of homeowners insurance sampled from the telephone yellow pages.

None of these sampling methodologies is perfect; each overlooks some unknown share of all transactions. And little is known about whether discrimination levels are higher or lower in these untested transactions. For example, analysis of the sampled homes advertised for sale and selected for testing in the national Housing Discrimination Study found that the vast majority of both houses and real estate offices were in predominantly white neighborhoods. As a result, houses for sale in minority or integrated neighborhoods were underrepresented in the national sample. And there are good reasons to believe that the incidence of discrimination against minority home buyers may be different in these neighborhoods than in predominantly white neighborhoods. Thus, the decision about how to sample transactions for testing has major implications for the interpretation of resulting discrimination measures. Ideally, we would like to sample from all transactions, or in a way that fully reflects the transactions minorities are likely to experience. But too little is known at this point about how people search for homes, apply for jobs, select a lending institution or insurance company, or choose a car dealership. Learning more about these market processes—and whether they differ for minorities and whites—should be an adjunct to development of testing methodologies.



Field Implementation Challenges. Fielding a nationwide paired testing study presents significant implementation challenges. Qualified testers must be recruited and trained in multiple sites. Random samples of transactions must be selected for testing. Testers must be closely supervised to ensure that they follow a single set of consistent procedures. And data from each test must be accurately recorded for future analysis. These demands require well-qualified field managers working on-site in each location where testing is being conducted, as well as centralized supervisory staff to ensure that the same procedures are being implemented properly in all of the sites.

In the two national studies of housing discrimination that HUD has sponsored, these implementation challenges have been addressed by sub-contracting with local fair housing organizations to recruit and train testers and to conduct the testing according to centrally prescribed procedures. Sampling was conducted centrally, and field supervisors provided oversight, coordination, and problem solving for all of the local agencies. This strategy has the advantage of relying upon established organizations with testing experience and a pool of potential tester candidates who know their local markets. Its potential disadvantages are that experienced testing organizations do not exist in all metro areas and that some established organizations may be unwilling or unable to adhere to a standardized set of testing procedures, particularly if these procedures differ from those normally followed locally. This approach may work less well outside the area of housing. The infrastructure of community-based organizations with experience conducting tests in the area of housing is far more developed than is the case in the area of employment or public accommodations. Further, in some instances results achieved by these groups might be perceived as biased.

An alternative to relying on local fair housing agencies that was used by the Urban Institute in its pilot tests of employment discrimination is to send central office staff into the field to recruit and train testers, draw samples, and supervise the testing. This strategy has the advantage of consistent procedures and quality control, and it does not assume that testing capacity exists in every location. However, because the testing managers do not know the local area, the challenges of recruiting testers and scheduling tester travel to and from assignments can be daunting. Further, this approach has never been used to conduct testing studies in more than three cities at the same time.

Balancing Research and Enforcement. The national report card can and should balance the interests of research and enforcement more effectively than has been done in past testing studies. Many researchers believe that testing designed primarily for measurement purposes must be carefully insulated from possible application for enforcement. They maintain that individual test results should never be used as evidence in litigation, and that the identities of both testers and tested institutions should be kept confidential. They argue that these restrictions protect the studies from potentially fatal defects:

- If testing organizations intend to use test results for enforcement purposes, they might target particular institutions or neighborhoods for testing rather than selecting a random sample, and they might record data in a form that



makes it useful legal evidence rather than in a form that makes it reliable quantitative data.

- If individual test results could be used as evidence, testers (or the testing organizations) might bring suits before all of the tests in a research sample had been completed. This would divulge the fact that testing was underway in a community and might change the behavior of the institutions and individuals being tested, invalidating all of the results for that site.
- If individual test results are used by civil rights organizations as evidence in litigation, people might believe that the testing was conducted by advocates, and that the methods were not objective or unbiased. This perception could undermine the credibility and persuasive power of the overall research findings.
- If a court rejected a discrimination claim that was based on evidence from one or more research tests, critics might try to discredit the report by disingenuously arguing that the study was wrong in all of the instances in which it found discrimination. The court's finding in one case, then, might "zero out" all of the research results.

These are all legitimate concerns, but it should be possible to address them effectively without completely prohibiting the use of research test results for enforcement purposes. Procedures can be developed that ensure objective measurement remains the primary purpose of the national report card, but that the test results can also be used for enforcement as a secondary objective. For example:

- Sampling procedures and tester reporting forms should be designed by researchers to ensure that the test results are statistically representative and that the data are suitable for quantitative analysis.
- Testers (and testing organizations) would be barred from bringing litigation until after all of the research tests had been completed.
- Local testing organizations would create a "firewall" between their investigative or enforcement activities and their research testing, so that the research testing is carried out according to research protocols with no influence from other activities of the organization.
- After all research tests were completed, individual test results could be used to identify institutions, locations, or sectors where discrimination appears prevalent, and to target additional, enforcement testing. If this follow-up testing resulted in litigation, the original test results would not be used as the primary testing evidence in the case.

These measures would protect the integrity of research testing without completely barring the use of individual test findings for enforcement purposes. This would avoid the quandary faced by some testing studies—that individual test results provide strong evidence of serious discrimination by a particular institution, that must be obscured and kept secret, even after the research has been completed and released. Finally, it should be noted that the potential for enforcement sanctions being levied against the former subjects of a research project raises new questions in human subjects ethics that have yet to be fully addressed.



Discrimination against Whom? The United States is becoming increasingly diverse, and African Americans are not the only racial or ethnic group to experience discrimination. Therefore, the national report card should not focus exclusively on discrimination against African Americans. However, not every ethnic group faces persistent barriers to opportunity and upward mobility. And it would not be feasible to measure the incidence of discrimination experienced by every racial or ethnic minority group in the U.S.¹⁴ Therefore, the report card should focus on discrimination against a limited number of racial and ethnic minorities, selected on the basis of evidence of past discrimination, persistent inequality, or institutional pressures which may lead to discrimination in the future. Likely candidates would include African Americans, Hispanic citizens and noncitizens, Asians, and Native Americans.¹⁵

Conclusion

In sum, a national report card would be valuable in an era in which patterns of inequality continue to follow racial and ethnic lines, public opinion is widely divided on the contribution of discrimination to these uneven outcomes, and the scope of affirmative action is being circumscribed. A report card could serve as a factual baseline for a national conversation on race, helping to avoid misguided policies that flow from premature claims of the advent of a colorblind society or unsupported victim-claiming. At the same time, the report card could help strategically target the enforcement efforts of civil rights agencies, and align them with the dictates of the Government Performance and Results Act. It could also help policymakers assess the degree to which discrimination might be serving as a barrier to the achievement of other policy goals, specifically reducing barriers to work for welfare recipients.

The authors of the following chapters in this volume argue persuasively that paired testing should play a central role in the development of the national report card. The accumulated experience of researchers and practitioners is sufficiently developed to design and implement national testing studies in housing and employment. National-level testing is also feasible in selected areas of sales (automobiles) and public accommodations (health clubs). Meanwhile experimentation should continue in the application of testing to other areas of economic life, including mortgage and commercial lending. Finally, testing results should be complemented by the development of longitudinal data in such areas as wage rates, home ownership, and minority firms' access to credit.

Endnotes

1. As discussed further below, people resist drawing a conclusion of discrimination based solely on evidence of disparities between minority and white outcomes or on evidence that relies on multivariate statistical techniques that are not readily understood.
2. The 1993 Government Performance and Results Act sets several goals, including: (1) holding federal agencies accountable for achieving program results; (2) helping federal agencies improve program delivery; and (3) improving congressional decisionmaking by providing more objective information on federal programs and spending.



3. Randomized tests would not be the only way that the civil rights enforcement agencies could determine the effectiveness of differing enforcement strategies. Follow-up examinations of firms previously found to be in violation would serve the same purpose.
4. The U.S. Department of Commerce's recently-announced benchmarks for retaining or phasing out price preferences for small, disadvantaged minority contractors represent a related attempt to systematically tie longitudinal measures of minority firm disadvantage to government policy. In this case, the Commerce Department will periodically assess the degree to which federal procurement funds are received by minority firms. Controlling, among other things, for size and government experience, price preferences will be extended in sectors and regions where the total share of government dollars that goes to minority firms is lower than their proportional representation among all firms. See the *New York Times* (1998).
5. It is important not to view these categories as immutable. The definition of "minority" groups is socially determined, differing across societies and subject to change over time.
6. In a recent article, Perlmann and Waldinger (1998, 73) wrote: "Having originated from anywhere but Europe, today's newcomers are visibly identifiable in a mainly white society still not cured of its racism. Moreover other changes in the structure of the U.S. economy aggravate discrimination's ill effects."
7. Testing studies are not unique in this regard. Statistical studies of discrimination can also shed light on variations in discriminatory treatment.
8. It is important to note that some researchers oppose this linked approach to research and enforcement testing, arguing that the long-term political viability and credibility of research testing requires a "firewall" between research and enforcement tests.
9. Note that some, but not all, paired testing studies use a "double blind" design, in which neither the white nor the minority tester knows what his or her partner experienced. This is difficult to achieve in testing efforts where partners need to know a lot about each others' experiences in order to ensure that their responses are matched.
10. Note that some testing managers argue strongly for the use of hidden tape recorders, particularly in an enforcement context. Tapes make it easier for testers to remember everything that happened in a complicated encounter, provide a mechanism for quality control, and can serve as valuable evidence in court. However, some states forbid the use of hidden tape recorders (even where one party to the conversation consents).
11. Federal law makes it illegal to provide false information on a credit application *with intent to defraud*. Some testing advocates argue that submitting false information as part of a paired test—when the tester will not actually borrow money or incur any other financial obligation—does not violate this law. The question has not yet been litigated, but some organizations may be willing to incur the risk.
12. Studies might also be conducted in non-metropolitan areas, with tests being carried out in a number of areas of economic activity. The results would generate suggestive, if anecdotal, results that could be built upon by broader initiatives, if warranted.
13. In each substantive area of investigation, the results of paired testing and analysis of aggregate data could be compared to enforcement data to determine how closely they are aligned.
14. This report card will focus exclusively on racial/ethnic discrimination, even though discrimination on the basis of gender, family composition, and disability status are serious issues as well.
15. No research testing conducted to date has focused on discrimination against Native Americans. Native American populations are highly concentrated in and around tribal lands and in a very small number of metropolitan areas. Thus, the sample of locations needed to generate national estimates of discrimination against Native Americans would be very different from samples designed to measure discrimination against other racial and ethnic minorities. See Kingsley, Mikelsons, and Herbig (1996).



References

- Ayres, Ian. 1991. "Fair Driving: Gender and Race Discrimination in Retail Car Negotiations." *Harvard Law Review* 104:817–72.
- Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *American Economic Review* 85:304–21.
- Bobo, Lawrence, et al. 1995. "Public Opinion Before and After the Spring of Discontent." In *The Los Angeles Riots: Lessons for the Urban Future*, edited by M. Baldassare. Boulder, Colo.: Westview Press.
- Cross, Harry et al. 1990. *Employer Hiring Practices: Differential Treatment of Hispanic and Anglo Job Seekers*. Washington, D.C.: The Urban Institute.
- Davis, James A., and Tom W. Smith. 1996. *General Social Surveys, 1972–1996*. Chicago: National Opinion Research Center.
- Enchautegui, Maria, Michael Fix, Pamela Loprest, Sarah von der Lippe, and Douglas Wissoker. 1996. *Do Minority-Owned Businesses Get a Fair Share of Government Contracts?* Washington, D.C.: Urban Institute Press.
- Farley, Reynolds. 1997a. "Average Hourly Earnings for Persons with Positive Earnings in 1996." Derived from U.S. Census Bureau, Annual Demographic File, March. Supplemental data from the Current Population Survey.
- _____. 1997b. "Per Capita Hours of Employment in 1996 for All Persons 25 to 59." Derived from U.S. Census Bureau, Annual Demographic File, March. Supplemental data from the Current Population Survey.
- Fix, Michael, and Jeffrey Passel. 1994. *Immigration and Immigrants, Setting the Record Straight*. Washington, D.C.: The Urban Institute.
- Hakken, Jon. 1979. *Discrimination against Chicanos in the Dallas Rental Housing Market: An Experimental Extension of the Housing Market Practices Survey*. Washington, D.C.: U.S. Department of Housing and Urban Development.
- James, Franklin, and Steve W. DelCastillo. 1992. *We May Be Making Progress toward Equal Access to Jobs: Evidence from Recent Audits*. University of Colorado at Denver.
- Kingsley, G. Thomas, Maris Mickelsons, and Carla Herbig. 1996. *Problems and Needs of American Indians and Alaska Natives*. A report prepared by the Urban Institute for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, D.C.: The Urban Institute.
- Massey, Douglas, and Nancy Denton. 1993. *American Apartheid*. Cambridge: Harvard University Press.
- Miller, Richard, and Austin Sarat. 1980. "The Emergence and Transformation of Disputes." *Law and Society Review* 15(3–4):525–65.



- New York Times*. 1998. "White House Revised Policy on Contracts for Minorities," June 25, A-1.
- Orfield, Gary, and Susan Eaton. 1996. *Dismantling Desegregation: The Quiet Reversal of Brown vs. Board of Education*. New York: The New Press.
- Patterson, Orlando. 1997. *The Ordeal of Integration: Progress and Resentment in America's "Racial" Crisis*. Washington, D.C.: Civitas/Counterpoint.
- Perlmann, Joel, and Roger Waldinger. 1998. "Are the Children of Today's Immigrants Making It?" *The Public Interest* Summer: 73–96.
- Ridley, Stanley, James A. Bayton, and Janice Hamilton Outtz. 1989. Taxi Service in the District of Columbia: Is It Influenced by the Patron's Race and Destination? Washington, D.C.: The Washington Lawyers' Committee for Civil Rights under the Law. Mimeographed.
- Ross, Stephen, and John Yinger. Forthcoming. "Does Discrimination Exist? The Boston Fed Study and Its Critics." In *Mortgage Lending Discrimination: A Review of the Evidence*, edited by Margery Austin Turner and Felicity Skidmore. Washington, D.C.: The Urban Institute.
- Turner, Margery, Michael Fix, and Raymond Struyk. 1991. *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring*. Washington, D.C. Urban Institute Press.
- Turner, Margery, Raymond Struyk, and John Yinger. 1991. *Housing Discrimination Study Synthesis*. Washington, D.C.: U.S. Department of Housing and Urban Development.
- U.S. General Accounting Office (GAO). 1990. *Immigration Reform: Employer Sanctions and the Question of Discrimination*. GAO/GGD-9062, March.
- Wienck, Ronald, et al. 1979. *Measuring Racial Discrimination in American Housing Markets: The Housing Practices Survey*. Washington, D.C.: U.S. Department of Housing and Urban Development.
- Wissoker, Douglas, Wendy Zimmermann, and George Galster. 1998. *Testing for Discrimination in Home Insurance*. Washington, D.C.: The Urban Institute.
- Yinger, John. 1991. *Lessons about Discrimination against Blacks and Hispanics from the Housing Discrimination Study*. Washington, D.C.: The Urban Institute.



Chapter 2

Testing for Discrimination in Housing and Related Markets

JOHN YINGER

Introduction

Testing, also called auditing, has transformed the way we look at racial and ethnic discrimination in housing and related markets.¹ Before testing came into widespread use, researchers studied housing discrimination primarily by looking for its impact on housing market outcomes, such as housing prices and rates of home ownership. Although this is an entirely respectable method of research, it involves abstract arguments and statistical principles that limit its currency in the debate about antidiscrimination policy. In contrast, testing provides a direct comparison of the treatment received by two equally qualified customers, one of whom belongs to a “protected class” as defined by our civil rights laws, and is therefore able to catch economic agents in the act of discriminating. As a result, this approach has a transparency and narrative power not found in previous research; it has proven to be invaluable in shedding light on discriminatory behavior.

This paper reviews testing-based research on discrimination in housing markets and makes recommendations for further research using this method. The paper begins with a brief history of housing testing and then discusses the strengths and weaknesses of the testing methodology. The following section presents a brief overview of the results from testing studies in housing, mortgage lending, and home insurance. The paper concludes by discussing a

comprehensive strategy for using testing to measure continuing discrimination in housing, and thereby to shed light on the progress of antidiscrimination programs.

A Brief History of Testing in Housing and Related Markets²

Fair housing tests were first developed by public and private fair housing agencies as a method for determining whether a complaint had validity. In the past, most activity by such groups began with a complaint, that is, with the appearance of a black or Hispanic person (or someone in another protected class) who claimed that he or she had been unfairly denied access to a house or an apartment. Agencies learned that they could establish whether a house or an apartment was indeed available and whether the complainant had indeed been unfairly denied access to it by sending a comparable white person to inquire about the same unit. When the white person was offered the unit that the black or Hispanic person was denied, the agency had powerful evidence of discrimination, for both administrative and legal purposes. By the early 1970s, many fair housing groups had experience bringing testing evidence into court, and several testing manuals were available.

Researchers then discovered that they could measure discrimination by conducting fair housing tests for a sample of housing agents or of advertised housing units. The first examples of testing research in the United States were small-scale studies conducted in Southern California in 1955 and 1971, followed by a large testing study in Detroit in 1974–75. A large testing study was also conducted in Great Britain in 1967.³ Testing became a highly visible research tool when, in 1977, the U.S. Department of Housing and Urban Development sponsored the Housing Market Practices Survey, which was a national study of housing discrimination against African Americans (Wienk et al. 1979). HMPS, as it came to be called, conducted 3,264 tests in 40 metropolitan areas and found evidence of significant discrimination against blacks in both the sales and rental markets. A follow-up testing study in Dallas found high levels of discrimination against Hispanics, particularly those with dark skin (Hakken 1979).

The pioneering HMPS report made it clear that fair housing tests are an appropriate and feasible method for studying discrimination in housing. Moreover, the strong HMPS results played a major role, albeit after a nine-year lag, in the passage of the 1988 amendments to the Fair Housing Act. It did not take long, therefore, for many fair housing agencies and researchers to conduct additional testing studies. Between 1977 and 1990, at least 72 other testing studies were conducted in individual cities. All these studies except Roychoudhury and Goodman (1992, 1996) are reviewed in Galster (1990a, 1990b). Virtually all of these studies provided further evidence of discrimination.

Thanks to the power of the HMPS results and the mounting evidence of continuing discrimination, the U.S. Department of Housing and Urban Development decided to sponsor a second national testing study. This study, called Housing Discrimination Study or HDS, conducted tests in 25 metropolitan areas in 1989. Several results from this study are reviewed in the next section.



During the past three years, fair housing groups have conducted black-white testing studies in five metropolitan areas (Fresno, Montgomery, New Orleans, San Antonio, and Washington, D.C.) and Hispanic-white testing studies in three areas (Fresno, San Antonio, and Washington, D.C.).⁴ These studies, which are also discussed below, used standard testing procedures but have not yet been scrutinized by scholars.

Strengths and Weaknesses of Testing

To understand the strengths and weakness of testing as a way to study discrimination, it is important to begin with a precise definition of discrimination. According to our fair housing laws, discrimination exists if an economic agent violates either one of two standards. The first standard involves the “disparate treatment” of customers on the basis of their membership in a protected class. According to this disparate-treatment standard, any economic agent who applies different rules to people in protected groups is practicing discrimination.

The second standard involves the use of practices with a “disparate” or “adverse impact” on the members of a protected class. Under the Fair Housing Act and the Equal Credit Opportunity Act, along with civil rights legislation that applies to employment and public accommodations, economic agents are discriminating whenever they use practices that do not explicitly consider a person’s group membership but instead have an adverse impact on a protected class without any “business necessity” (Schwemm 1992; Vartanian, Ledig, and Babitz 1995). Under current law, if a particular practice can be shown to have an adverse impact on a protected class, then the burden of proof shifts to the business, which must then support a business necessity claim. This so-called “effects test” rules out a potential loophole in an antidiscrimination law, because it prevents a business from disguising its mistreatment of a racial or ethnic group as an apparently neutral policy based on a characteristic that is highly correlated with race or ethnicity but not necessary for business success. Moreover, this test requires all firms to eliminate outmoded rules of thumb and other unnecessary business practices that have a disproportionate impact on a protected class.

Testing provides an ideal way to observe and measure discrimination as defined by the disparate-treatment standard. To be more specific, testing has four principal advantages as a tool for measuring disparate-treatment discrimination:

First, testing makes it possible for researchers to compare the treatment of two people, one of whom is in a protected class, who are equally qualified for renting or buying housing (or for some other transaction) and who encounter equivalent circumstances in the marketplace. In particular, researchers can match similar individuals; give them the same training; assign them similar or identical characteristics, such as income, for the purposes of the test; and send them to visit economic agents within a short time of each other. These procedures are designed to ensure that testing teammates do not differ significantly on any characteristic, other than race, ethnicity, or sex, that is relevant to their treatment by economic agents, so that any observed differences can be attrib-



uted to discrimination. In technical terms, the testing design makes it possible to minimize, if not rule out altogether, the possibility that differences in treatment are caused by variables that the researcher cannot observe. This problem, and the resulting “omitted variable bias,” is, of course, one of the principal threats to the validity of inferences about discrimination from the other leading method for studying discrimination, namely multiple regression.⁵

It is worth mentioning that testing studies may not have a great advantage over regression studies in controlling for variables that can be observed, such as income or family size, so long as the regression study has many observations to work with.⁶ A testing study’s advantage with these variables comes from the fact that, unlike a regression study, it does not need to make an assumption about the functional form of the relationship between control variables and agent behavior. However, specification may not be a central issue, and regression studies can now deal with it in a very general way through the use of various specification tests. Nevertheless, testing studies have a great advantage over regression studies in controlling for variables that do not appear in the types of data sets that regression studies use, such as those derived from the U.S. Census or the American Housing Survey. To be specific, testing studies, unlike regression studies, can control for the types of queries customers make, for the way customers behave when talking with an economic agent, and for the timing of visits to these agents.⁷

Second, a testing study can observe many types of agent behavior and therefore can determine if different agents discriminate in different ways. Existing evidence from testing studies reveals discrimination in an amazingly wide range of agent behavior; no other method could begin to pick this up. Regression studies, for example, almost invariably concentrate on a single type of agent behavior, such as a loan approval decision, or a single housing market outcome, such as the home ownership rate, and therefore cannot shed light on the complexity of discriminatory behavior.

Third, a testing study has great narrative power. Anyone can imagine what it would be like to be treated the way testers in a protected class are treated, and anyone can understand why differences in treatment between equally qualified testing teammates constitute discrimination. This narrative power and plausibility cannot be matched by a regression study, which depends for its credibility on abstract arguments about data quality, omitted variable bias, and interpretation.

Fourth, testing studies provide a unique opportunity to observe the circumstances under which discrimination occurs. This feature makes it possible to test hypotheses about the causes of discrimination and to improve the effectiveness of fair housing enforcement. I will return to these topics later.

Testing also has several disadvantages. First, a testing study does not provide evidence on discrimination in general but instead provides evidence on discrimination in the realm defined by the study’s sampling frame. For example, HDS measures discrimination that qualified blacks and Hispanics could expect to encounter if they inquired about housing advertised in a major metropolitan newspaper (see Yinger 1995). HDS does not reveal the discrimination experienced by the average black or Hispanic household, both because it does not observe housing that is not advertised in a major newspaper and



because it is not based on the socioeconomic characteristics of the average black or Hispanic household. A regression study of home ownership rates based on national data could, in principle, come closer to providing this type of measure. However, a testing study can observe the extent to which discrimination varies with the characteristics of people in protected classes (or of the housing they seek). (See Yinger 1995.)

Second, some types of agent behavior are difficult to observe with the testing methodology. For the purposes of this paper, the main example of this limitation is that it is difficult to design a testing procedure for measuring discrimination in loan approval. Not only would such a procedure be complicated (because it would involve all the information involved in a credit check), but it might run into laws against providing false information on credit applications. I will return to this issue in a later section. Another example comes from the recent behavior of some landlords attempting to make it difficult for testing to uncover their discriminatory behavior. These landlords refuse to show available apartments to any potential tenant until that tenant fills out a detailed application. As a result, discrimination in showing or renting apartments cannot be uncovered without a relatively complicated testing procedure, which may involve credit information.

Third, testing has not yet been used to study disparate-impact discrimination, and it does not appear to be well suited for such an effort. If a particular type of disparate-impact discrimination could be identified, then, in principle, a test could be designed to see if it exists. For example, if landlords reject people from certain occupations, even though their occupational status has nothing to do with their desirability as tenants, and members of a protected class are relatively concentrated in these occupations, then an occupation-based testing study could be conducted.⁸ However, we know so little about disparate-impact discrimination that it would be difficult to identify the relevant variables ahead of time. Moreover, different housing agents may use different variables to practice disparate-impact discrimination, so a testing study, which by definition must be based on a single variable, might miss most of the disparate-impact discrimination that is taking place.

Finally, a testing study provides compelling, easy-to-understand approximations of the extent of discrimination, but it cannot easily provide a precise measure of discrimination, even by the disparate-treatment standard.⁹ Consider first the *incidence* of discrimination for some type of agent behavior, such as showing an advertised apartment. The vast majority of testing studies measure this incidence using one of two simple measures. The gross incidence of unfavorable treatment, called the “gross measure,” is the share of tests in which the tester in the protected class is treated less favorably than his or her teammate. As pointed out by Wienk et al. (1979), however, discrimination is the systematic unfavorable treatment of a protected class, so a measure of discrimination should not be affected by random differences in treatment. It should not reflect, for example, a case in which an apartment is rented after a white tester sees it but before her black teammate even inquires about it. Wienk et al. argue that the share of tests in which the teammate from the protected class is favored provides an estimate of the extent to which random factors are at work. Thus, they suggest looking at the net incidence of unfavorable treatment, called the



“net measure,” which equals the gross measure minus the share of tests in which the tester in the protected class is treated more favorably.

The problem, of course, is that the net and gross measures are often far apart. Some economists favor the net measure because it is conservative; by ignoring random factors, the gross measure may overstate the incidence of discrimination, whereas the net measure can be seen as a lower bound. As pointed out by Fix, Galster, and Struyk (1993), and Yinger (1993), however, this lower bound can be quite inaccurate, because some actions that favor members of a protected class are the result of systematic, not random factors. Consider, for example, the case in which real estate brokers decide not to show houses in largely black neighborhoods to white customers; the net measure implicitly regards this behavior as random, when in fact it is a type of systematic behavior that in no way offsets discrimination that blacks may encounter when they look at houses in white neighborhoods.

An alternative approach taken by Ondrich, Ross, and Yinger (1997) is to use the features of a testing study to estimate the roles played by systematic and random factors and then to explicitly remove random factors when calculating the incidence of discrimination. For example, they estimate the extent to which unfavorable treatment varies with a test’s circumstances and then identify cases in which white testers are favored for systematic reasons—cases that should not be “netted out” in calculating discrimination. This approach reveals that the net measure is not really a measure of the incidence of discrimination; instead, it measures a more conservative concept, namely, the extent to which members of the protected class are more likely to encounter unfavorable treatment than are whites (or white men).

For types of agent behavior that are a matter of degree, such as the number of houses shown to a customer, tests have also been used to explore the *severity* of discrimination, defined as the extent to which some people experience less favorable treatment solely because of their membership in a protected class. All existing studies employ a net measure of severity or something analogous to it, defined as the average difference in treatment between white and minority testers. However, the above analysis of the incidence of discrimination implies that this approach understates discrimination because it nets out cases in which members of a protected class are favored for systematic reasons.¹⁰ Formally, the literature estimates the magnitude of the difference in treatment between whites and members of various protected classes, not of discrimination.

Results of Testing in Housing and Related Markets

Testing has been used to observe discrimination in the marketing behavior of landlords and real estate brokers, in the preapplication behavior of lenders, and in the marketing behavior of home insurance agents. This section reviews some key results from this research.

The Housing Market¹¹

The 1989 national study, HDS, examined discrimination against both blacks and Hispanics. Black-white tests were conducted in 20 metropolitan areas and



Hispanic-white tests were conducted in 13 areas, with both types of test in eight of these areas. The sites were selected so that the tests would yield nationally representative results. In each site, housing advertisements were randomly selected from the major metropolitan newspaper for several weekends during the spring and summer of 1989. Each test was based on one of these advertisements. Testers were all given the same training, test teammates were assigned virtually identical economic and family characteristics for the purposes of each test, and the order in which teammates conducted each test was determined randomly. In total, HDS conducted 1,081 black-white tests in the sales market, 801 black-white tests in the rental market, 1,076 Hispanic-white tests in the sales market, and 787 Hispanic-white tests in the rental market.

HDS examined discrimination in a wide range of housing agent behavior, and table 1 presents a few of the incidence results.¹² For example, this table reveals that, based on the net measure, black renters faced a 10.7 percent chance of being excluded altogether from housing made available to comparable white renters and a 23.5 percent chance of learning about fewer apartments. Moreover, these estimates of discrimination are statistically significant for virtually every type of agent behavior.

Table 1 also reveals that the gross measure can be close to the net measure or far above it. To lessen the resultant uncertainty about the true incidence of discrimination, it is possible, as mentioned earlier, to separate discrimination from random differences in treatment using statistical procedures.¹³ Ondrich, Ross, and Yinger, for example, apply this approach to the HDS data and obtain esti-

Table 1 *The Incidence of Discrimination in Housing, 1989 Housing Discrimination Study*

	Black-white audits		Hispanic-white audits	
	Net	Gross	Net	Gross
Sales Audits				
Excluded from Available Units	6.3*	7.6	4.5+	7.5
Advertised Unit Inspected	5.6*	13.3	4.2*	13.2
Number of Houses Made Available	19.4*	44.1	16.5*	43.6
Auditor Asked to Call Back	3.3*	25.9	11.5*	30.4
Auditor Received Follow-up Call	7.7*	18.5	5.5*	16.4
Auditor Received Positive Comments on House	12.5*	47.9	7.5*	47.5
Agent Offered to Help Auditor Find Financing	11.3*	24.4	4.4+	22.1
Rental Audits				
Excluded from Available Units	10.7*	15.1	6.5*	12.1
Advertised Unit Inspected	12.5*	23.0	5.1	17.6
Number of Apartments Made Available	23.3*	41.4	9.8	34.6
Auditor Asked to Call Back	15.8*	30.5	8.6*	28.5
Auditor Received Special Rental Incentives	5.4*	10.3	5.1*	12.6
Auditor Received Positive Comments on Apartment	16.8*	48.4	14.6*	46.4

Note: A * indicates statistical significance at the 5 percent two-tailed level based on a fixed-effects logit procedure (net measure only). A + indicates statistical significance at the 5 percent one-tailed level based on a fixed-effects logit procedure (net measure only).

Source: Yinger (1998).



mates of discrimination that are always above the net measure, but are closer to the net measure in some cases and closer to the gross measure in others. Because of the HDS tests' design, Ondrich, Ross, and Yinger (1997) cannot provide an exact measure of the incidence of discrimination, but they can calculate lower and upper bounds. For several types of behavior, the lower bound for their measure of discrimination is considerably above the simple net measure. For "advertised unit inspected," for example, their lower bound is 8.2 percent, which is considerably above the net measure for sales audits in table 1 (5.6 percent). For "auditor asked to call back," the difference is even greater: 15.4 percent for the lower bound compared to the 3.3 percent net measure in table 1. The simple gross measure falls below the Ondrich/Ross/Yinger upper bound in a few cases, but is close to the parametric measure in others. Again, for "auditor asked to call back," the gross measure in table 1 is 25.9 percent, whereas their upper bound is 36.7 percent.¹⁴

Another approach to the incidence of discrimination is provided by Ondrich, Stricker, and Yinger (1998). This approach, which could be seen as a compromise between the simple net measure and the complex Ondrich/Ross/Yinger measure, uses a well-known statistical procedure called logit analysis to adjust the net measure for the observable differences between teammates, such as in age or order of visit. This approach yields an approximate incidence of discrimination based on one of two assumptions. The first is that discrimination takes the form of a fixed proportional difference between the probability that a favorable action will be taken for white customers and the probability it will be taken for customers in a protected class. The second is that discrimination takes the form of a fixed absolute difference between these two probabilities.

Application of this approach to the HDS data yields the results in tables 2 and 3. In every case, the logit-based estimates of the incidence of discrimination in columns four and five are larger than the simple net measure in the third column.¹⁵ For the three housing-availability variables at the bottom of these tables, the incidence of discrimination is at least 10 percent, and perhaps as high as 40 percent, for both blacks and Hispanics. Moreover, real estate brokers also are much more likely to offer financial assistance to black than to white customers.

The incidence of discrimination does not appear to be abating in recent years. A detailed comparison of the HDS results with those of the 1977 national study finds no clear evidence of a trend in either direction (Yinger 1995). The preliminary evidence that has become available since HDS shows no sign of a trend since 1989, either. HDS calculated a summary index of discrimination across many types of agent behavior, including the provision of information about available housing, agent efforts to help complete a housing transaction, information about financing or credit checks, and steering toward certain types of neighborhoods (see Turner, Struyk, and Yinger 1991). This index indicated that the probability of some form of discrimination (using a gross measure) was at about 50 percent for both blacks and Hispanics in both the rental and sales markets. On the basis of a similar index, the five studies conducted in the 1990s find that the gross measure of discrimination in rental housing is at least 50 percent (and as high as 77 percent) against both blacks and Hispanics in the first four areas and about 40 percent against blacks and Hispanics in the sales and rental markets in the Washington, D.C., area.¹⁶ See Fair Housing



Table 2 *Approximations of the Probability of Discrimination, Black Audits, 1989 Housing Discrimination Study*

Broker Action	Probability of Action for		Difference in Probability of Action	Odds Ratio	Probability Measure of Discrimination		Maximum Possible Discrimination, Absolute Gap Measure ^d
	White ^a	Black			Fixed Percentage Gap ^b	Fixed Absolute Gap ^c	
Call Back	0.475	0.437	0.038	1.204	0.097	0.046	0.046
Ask About Income	0.222	0.303	-0.081	0.411	0.847	0.188	0.219
Follow-Up Call Made	0.267	0.174	0.093	1.926	0.404	0.108	0.162
Ask About Needs	0.751	0.686	0.065	1.374	0.085	0.064	0.079
Financial Assistance Offered	0.376	0.263	0.113	2.025	0.390	0.147	0.175
Advertised Unit Inspected	0.630	0.573	0.057	1.834	0.236	0.149	0.151
Similar Unit Inspected	0.350	0.259	0.091	1.994	0.392	0.137	0.171
Advertised Unit Available	0.886	0.810	0.076	2.363	0.154	0.134	0.212

^aShare of audits in which the white customer received treatment.

^bEstimated fixed proportion by which probability of action for black customers falls short of probability of action for whites.

^cEstimated fixed amount by which probability of action for black customers falls short of probability of action for whites.

^dMaximum absolute gap between the probability of action for white and black customers, given the estimated odds ratio.

Source: Ondrich, Stricker, and Yinger (1998).

Council of Fresno County (1995), Central Alabama Fair Housing Center (1996), Fair Housing Action Center (1996), The Fair Housing Council of Greater Washington (1997), and San Antonio Fair Housing Council (1997).¹⁷

Another form of index is provided by Yinger (1991). This index counts the number of times a black or Hispanic tester is treated less favorably on a single type of agent behavior than is his or her white teammate and then subtracts

Table 3 *Approximations of the Probability of Discrimination, Hispanic Audits, 1989 Housing Discrimination Study*

Broker Action	Probability of Action for		Difference in Probability of Action	Odds Ratio	Probability Measure of Discrimination		Maximum Possible Discrimination, Absolute Gap Measure ^d
	White ^a	Hispanic			Fixed Percentage Gap ^b	Fixed Absolute Gap ^c	
Call Back	0.490	0.372	0.118	1.614	0.239	0.117	0.119
Ask About Income	0.294	0.309	0.015	1.005	0.003	0.001	0.001
Follow-Up Call Made	0.248	0.191	0.057	3.811	0.579	0.168	0.323
Ask About Needs	0.767	0.758	0.009	1.413	0.088	0.067	0.086
Financial Assistance Offered	0.377	0.333	0.044	1.358	0.182	0.069	0.076
Advertised Unit Inspected	0.667	0.613	0.054	2.190	0.284	0.189	0.193
Similar Unit Inspected	0.355	0.292	0.063	2.168	0.430	0.153	0.191
Advertised Unit Available	0.887	0.849	0.038	4.406	0.278	0.246	0.355

^aShare of audits in which the white customer received treatment.

^bEstimated fixed proportion by which probability of action for Hispanic customers falls short of probability of action for whites.

^cEstimated fixed amount by which probability of action for Hispanic customers falls short of probability of action for whites.

^dMaximum absolute gap between the probability of action for white and Hispanic customers, given the estimated odds ratio.

Source: Ondrich, Stricker, and Yinger (1998).



the number of times the white teammate is favored. The result is a “net” measure of the average number of acts of discrimination a black or Hispanic customer can expect to encounter during each visit to a housing agent. Application of this approach to the HDS data reveals that, on average, black and Hispanic home buyers can each expect to encounter about one act of discrimination each time they visit a real estate broker. The comparable estimates are 0.83 acts for black renters and 0.58 acts for Hispanic renters. Most of this discrimination involves housing availability. For black home buyers and renters and for Hispanic home buyers, the number of acts of discrimination involving housing availability alone falls between 0.35 and 0.40 for each visit to an agent. The comparable estimate is 0.23 acts for Hispanic renters. All of these estimates are highly significant statistically.

Several testing studies also examine the severity of discrimination in housing (Page 1995; Roychoudhry and Goodman 1992, 1996; Yinger 1986, 1993, 1995). According to the HDS data, for example, black home buyers learn about 23.7 percent fewer houses than do their white teammates, black renters learn about 24.5 percent fewer apartments, Hispanics learn about 25.6 percent fewer houses, and Hispanic renters learn about 10.9 percent fewer apartments (Yinger 1995, table 3.2). All of these net differences are statistically significant.

Overall, this research demonstrates that black and Hispanic home seekers continue to encounter discrimination in many aspects of a housing transaction. They are told about fewer available units and must put forth considerably more effort to obtain information and to complete a transaction. These barriers are not absolute, but they impose significant costs on black and Hispanic home seekers relative to comparable whites in the form of higher search costs, poorer housing outcomes, or both (Yinger 1995, ch. 6).

Mortgage Lending

The vast majority of home buyers require a mortgage loan. Thus, discrimination in mortgage lending can pose a major barrier to home ownership. The first applications of the testing methodology in mortgage lending were small pilot studies of lenders’ pre-application behavior in the 1980s and early 1990s. See Smith and Cloud (1996) and Lawton (1996). In 1993, a national testing study of lenders’ pre-application behavior was sponsored by the Department of Housing and Urban Development (HUD) and conducted by the National Fair Housing Alliance (NFHA), a coalition of local fair housing organizations. This study conducted black-white and Hispanic-white tests in several cities, with a large number of tests in Chicago and Oakland and a smaller number in Atlanta, Dallas, Denver, Detroit, Richmond, and Norfolk (see Smith and Cloud 1996). This study found adverse treatment of potential black and Hispanic borrowers in a wide range of lender behavior, including the following:

- requiring a credit check, completion of an application, or presentation of other documentation prior to scheduling an appointment with African-American and Latino persons while not imposing the same preappointment requirements on white persons;
- quoting more restrictive qualification standards and ratios to African-American and Latino persons;



- making exceptions to qualification standards for white persons while not making the same level of exceptions for African-American and Latino persons;
- requiring higher levels of escrow and reserve payments for African-American and Latino persons at the time of closing than for white persons;
- providing constructive advice (i.e., paying down debt, obtaining gift letters, explaining credit flaws) to white persons on how to circumvent a potential barrier to qualifications while not providing the same advice or quality of advice to African-American and Latino persons (Smith and Cloud 1996, pp. 598–99).

Thus, testing has proven to be a powerful tool for uncovering discriminatory practices by lenders in the preapplication or loan origination phase of a lending transaction. Fair housing groups also have conducted a few tests that cover mortgage applications (Smith and Cloud 1996), but no large-scale testing study of this type has yet been attempted.

Home Insurance

Home insurance is a requirement for obtaining a mortgage; if a household is denied home insurance, either because it belongs to a protected class or because it wants a house in a neighborhood where a protected class is concentrated, it may not be able to obtain a mortgage and therefore may not be able to buy a house at all. Differential treatment on the basis of a property's location is called redlining. A few testing studies suggest that discrimination and redlining occur in the market for home insurance and may even be common. In a pilot testing study of redlining in home insurance in Milwaukee, Squires and Velez (1988), found that insurance agents stated more stringent inspection standards for houses in black than in white neighborhoods. NFHA began supporting home insurance testing and in 1991 received a grant from HUD to conduct such tests in nine cities: Akron, Atlanta, Chicago, Cincinnati, Los Angeles, Louisville, Memphis, Milwaukee, and Toledo. This testing focused on insurance discrimination and redlining in the pre-application behavior of three large insurance companies. The tests were conducted over the telephone, so the testers were limited to people who had voices that revealed their ethnic identity. For more details, see Smith and Cloud (1997).

Each of the NFHA tests involved a match between two houses, two neighborhoods, and two testers. First, two houses were matched on the basis of type of construction, age, size, number of stories, number of rooms, type of basement, electrical system, locks, and fire detectors. Second, these two houses were in similar neighborhoods. None of the houses were near abandoned buildings or other hazards and matched houses were a similar distance from fire hydrants and fire stations. Most of the houses were in a working- or middle-class neighborhood.¹⁸ Third, testers were all assigned an identity in which they had good credit, had not filed for bankruptcy, and had not filed a home insurance claim or had a home insurance policy canceled during the last five years. Two types of tests were conducted. The first type looked for a combination of discrimination and redlining by matching a black (or Hispanic) tester who was



buying a house in a black (or Hispanic) neighborhood with a white tester who was buying a house in a white neighborhood.¹⁹ The second type looked for discrimination by matching a black (or Hispanic) tester with a white tester, both of whom were trying to buy a house in a white neighborhood.

This study found, in every city, high rates both of redlining against black and Hispanic neighborhoods and of discrimination against black and Hispanic applicants. Unfortunately, however, Smith and Cloud (1997) do not report separate incidence measures for discrimination and redlining nor do they indicate how many tests of each type were conducted. In terms of redlining, the study found that, compared with white applicants in white neighborhoods, black (or Hispanic) applicants in black (or Hispanic) neighborhoods were more likely to be quoted a higher price for coverage (per square foot or per dollar of replacement value); more likely to be offered less generous types of policies; less likely to receive a written quote that had been promised on the phone; more likely to be told about a company policy that made it impossible for the agent to issue the requested policy; and less likely to be told how to get around a restrictive company policy. The same types of difference in treatment arose in white neighborhoods between white testers and black or Hispanic testers.

The NFHA study was limited to three large insurance companies. To obtain information about insurance redlining in general, HUD funded another two-city testing study for the home insurance market in general. This study was not released in time to be reviewed in this chapter. It should be noted, however, that this study found little evidence of insurance redlining.

Monitoring Discrimination in Housing and Related Markets

In the interests of good management, political support, and effective design, the performance of antidiscrimination programs, like the progress of other programs, should be monitored. In other words, the federal government would be well served by a program that kept track of the extent of discrimination over time and attempted to determine the impact of enforcement activities on discrimination. This section discusses the role of testing in this type of monitoring program.

The Housing Market

Testing has no close competitors as a method for establishing the extent to which certain groups continue to face disparate-treatment discrimination in the sale and rental of housing. As a result, it must remain a key tool for monitoring the performance of the nation's antidiscrimination efforts in housing markets. This monitoring potential has barely been tapped, as we have only two national testing studies and no national testing-based measure of housing discrimination since the 1988 amendments to the Fair Housing Act were implemented.²⁰ Any serious antidiscrimination program should include regular testing in both the sales and rental housing markets.²¹

Testing is valuable not only because it yields measures of the incidence and severity of discrimination, but also because it provides insight into the



circumstances under which discrimination occurs. Several studies have found, for example, that discrimination in the sales market is often particularly high in integrated neighborhoods (see Ondrich, Stricker, and Yinger 1998; Page 1995; Yinger 1986, 1995).

This type of evidence is valuable for two reasons. First, it sheds light on the causes of discrimination. Several scholars have argued, for example, that real estate brokers may have an incentive to prevent racial or ethnic “tipping,” and may therefore discriminate heavily in integrated neighborhoods that are close to a tipping point. The evidence just cited supports this hypothesis. For a more detailed discussion of the causes of discrimination, see Yinger (1995).

Second, this type of evidence can help to guide enforcement activities. If testing studies can identify the types of neighborhoods in which discrimination is most likely to occur, for example, then the effectiveness of enforcement efforts can be enhanced by shifting some enforcement activities toward those neighborhoods.²²

The ability of existing testing studies to pursue these issues has been limited by the nature of the studies. For example, HMPS and HDS were both based on a random sample of advertisements in major metropolitan newspapers. Because houses in largely black and largely Hispanic neighborhoods are rarely advertised in major newspapers, those neighborhoods were under-represented in the resulting samples. Because hypotheses about the causes of discrimination rely heavily on changes in the incentives brokers and landlords face in different types of neighborhoods, the limited variation in neighborhood racial and ethnic composition limits the ability of scholars to untangle the causes of discrimination. Some progress on this topic has been made, but more progress may depend on new data based on a stratified sampling technique that ensures adequate representation of largely black and largely Hispanic neighborhoods.

The lack of variation in neighborhood racial and ethnic composition also limits the ability of testing studies to uncover racial and ethnic steering. Analysis of the HDS data found that under certain circumstances, real estate brokers were almost certain to steer, but that the low representation of houses in black and Hispanic neighborhoods meant that the opportunity to steer did not usually arise. For example, Yinger (1995) finds that if the advertised house is in a black neighborhood and the agent has access to additional houses in white neighborhoods through a multiple listing service, there is an 80 percent chance that those extra houses will be shown only to the white customer. Galster (1990b) obtains a similar result. On average, however, the incidence of steering and the differences in neighborhood racial composition when steering does occur are both quite low (Turner and Mickelsons 1992). Tests based on stratified samples also could shed more light on steering.

This recommendation to use stratified samples has one caveat. Two studies (Newburger 1995; Turner 1992) have found that houses in largely black neighborhoods not only are rarely advertised in the newspaper but also are rarely marketed with open houses, a technique that is widely used in largely white neighborhoods. A stratified sample can draw more advertised units in black neighborhoods, but the experience of people who inquire about these units may not be typical of the experience of people who obtain information about housing through the as-yet-unidentified informal mechanisms through which



houses in black neighborhoods are sold. Further research on the marketing of housing in largely black and Hispanic neighborhoods would be valuable and might eventually inform the selection of a sampling frame for a testing study.

Testing has proven to be an effective tool for measuring discrimination in both the sales and rental markets. As noted earlier, however, some rental housing market practices pose difficult challenges for testing studies. In particular, if a rental agent declines to show apartments to any applicants until they have completed an application, complete with credit information, then standard testing procedures cannot observe discrimination by this agent. To the extent that this practice has become widespread, perhaps as a way for landlords to protect themselves from testing, then new testing methods will have to be developed. If the application forms ask for a limited amount of information and are evaluated by landlords without formal credit checks, then tests can be extended simply by assigning somewhat more complete identities to testers. If, on the other hand, application forms required detailed information or landlords complete formal credit checks before showing apartments to prospective tenants, then testing may require the use of actual credit histories. This step would greatly complicate the design of a testing study but might be feasible. Because the same step is required to apply the testing methodology to loan approval decisions, it will be discussed more fully in the discussion of testing in mortgage markets.

Mortgage Lending

The 1993 testing study of the preapplication behavior of lenders, which was described earlier, provided valuable information about lending discrimination. Periodic replication of this type of study would be valuable.

A more difficult question is whether to attempt a testing study of the loan approval decision. Because it is illegal to falsify information on a credit application, a testing study of this type probably could not make use of false identities, but instead would have to use the actual identities of testers.²³ As a result, a testing study of loan approvals would have to rely on some combination of a large pool of potential testers, so that matches of actual financial characteristics could be found, and statistical methods to control for differences in teammates' credit histories. As explained earlier, a lack of complete matching may not be a serious problem; for variables that can be observed and measured (including the variables in an applicant's credit history), statistical controls may be almost as good as testing-based controls (see Ross 1997). If underwriting standards vary widely across lenders, however, a regression analysis may not be able to distinguish between disparate-treatment discrimination (using less favorable standards for members of a protected class), disparate-impact discrimination (using unjustified standards with a disparate impact on a protected class), and variation in underwriting standards with a legitimate business purpose.

Moreover, the management challenges confronting a hybrid testing/statistical study of discrimination in mortgage loan approvals are formidable.²⁴ First, most loan applicants are referred by real estate brokers, so the credibility of the testers could be enhanced by enlisting the cooperation of real estate brokers.



This step was conducted in a testing study of preapplication lender behavior in Philadelphia (see Lawton 1996). Second, detailed procedures for observing testers' actual credit histories would have to be developed. Testers with serious credit problems would have to be dropped. Information would have to be collected on the wide range of factors considered by lenders in making a loan-approval decision. Third, a large number of testers would have to be identified and, if possible, matched on actual credit characteristics. A large number of testers would also have to be trained. Because lenders would conduct an actual credit check on each tester, the study could not expect a tester to conduct many tests. In fact, it might not be possible for a tester to conduct more than one test because a person's credit records indicate whether a credit check was recently completed; lenders might become suspicious if an applicant had just experienced a credit check for another lender.²⁵ Consequently, a study of this type might require a veritable army of testers.

Given the magnitude of these challenges, does a pilot testing/statistical study of loan approvals make sense? I believe the answer to this question is affirmative, largely because existing studies of loan approvals, which do not use testing at all, are so controversial. A review of this controversy is beyond the scope of this paper (but see Goering and Wienk 1996; Yinger 1995). Suffice it to say that a major recent study, Munnell et al. (1996), which was published in *The American Economic Review*, the leading journal in the economic profession and which used the best data and best techniques ever applied to the topic, has been bitterly attacked by dozens of scholars who simply do not believe the study's conclusion that discrimination exists. I believe these attacks are unwarranted, but I must acknowledge that many scholars reject the study's findings, and indeed appear unwilling to accept any evidence based on the regression methodology. In addition, testing may be the only method that can isolate disparate-treatment discrimination. Given the importance of mortgage lending for access to home ownership and the power of the evidence that discrimination exists, both in preapplication behavior and in loan approval, there is a pressing need for a study that could provide less controversial evidence. I believe a carefully conducted testing/statistical study might fill this gap, and recommend that a small-scale study of this type, say in one metropolitan area, be conducted to determine its feasibility.

Home Insurance

Existing studies reveal that home insurance testing is challenging, far more challenging than testing in the sales or rental housing markets or testing the preapplication behavior of lenders, because it involves matching houses, neighborhoods, and testers. Moreover, insurance agents in different companies now often share computer databases, a practice that makes it difficult for a testing study to avoid detection. Nevertheless, the NFHA study provides striking evidence of discrimination and redlining in home insurance, and further small-scale testing studies in this market, building on the experience of the just-released study of a random sample of agencies (Wissoker et al. 1998), would no doubt pay dividends.



Conclusion

Testing is unmatched as a method for measuring disparate-treatment discrimination in housing markets, and it should continue to be the foundation of any effort to keep track of this discrimination over time. New testing methods may be required, however, to understand the full extent of racial steering, to uncover discrimination by landlords who show housing only to prospective tenants who have completed an application form, or to measure disparate-impact discrimination. Testing also has unique advantages as a method to shed light on the causes of discrimination and to design cost-effective antidiscrimination enforcement mechanisms—without sacrificing its power as a measurement tool. Its potential contributions to these topics has barely been tapped, however, and further research would be valuable.

Testing also has made significant contributions to an understanding of ethnic discrimination in preapplication behavior by mortgage lenders and of redlining in the provision of home insurance. Given the power of the testing methodology and the need for better information about discrimination in these markets, small-scale applications of testing to new types of behavior are also likely to be worthwhile. One possibility that is, in my view, particularly promising, is a study of discrimination in loan approval that makes use of testers' actual credit histories. This type of study would face several complex methodological hurdles, because it would have to collect and match information on testers' actual credit histories, it would need a large number of testers, and it might require the use of statistical controls. Nevertheless, I believe these hurdles could be overcome.

Overall, the most urgent need, in my view, is for another national testing study of discrimination in the sale and rental of housing. A study of this kind would not only reveal whether discrimination has declined since the implementation of the 1988 Fair Housing Amendments Act, but also provide a valuable opportunity to expand our knowledge about discrimination in integrated neighborhoods and to increase our understanding of the causes of discrimination. Some policymakers apparently agree with this view, and HUD recently announced the appropriation of money for just such a study.

Endnotes

1. Many people distinguish between “testing” done for enforcement purposes and “auditing” done for measurement or research purposes. This paper uses these two terms as synonyms (as does the title of the meeting at which this chapter was first presented).
2. This section draws heavily on Yinger (1995, ch. 2).
3. A more detailed discussion of these studies, along with citations, can be found in Yinger (1995, ch. 2).
4. In this paper, “Hispanic-white” is shorthand for Hispanic compared to non-Hispanic white.
5. For a more detailed discussion of this and other weaknesses of the regression methodology, see Yinger (1998).
6. The discussion of this issue here and below draws on Ross (1997) and on conversations with Stephen Ross.



7. For an example of the interpretation problems that can arise in a regression study of discrimination in car sales because of the lack of controls for customer behavior, see Yinger (1998).
8. Another example comes from the market for home insurance, where some agencies have policies to deny coverage to houses above a certain age or below a certain value. These policies have a disparate impact on blacks and Hispanics, whose houses are, on average, older and less valuable. As it turns out, these policies also involve disparate treatment: agencies are more likely to inform black and Hispanic customers than white customers about these policies and less likely to tell black and Hispanic customers how to get around them. See Smith and Cloud (1997).
9. The remaining paragraphs in this section draw heavily on Yinger (1998).
10. The severity and incidence of discrimination are logically connected. See Yinger (1993).
11. From its start through the discussion of table 1, this section draws heavily on Yinger (1998).
12. HDS researchers also examined the efforts of real estate brokers to “steer” black and Hispanic customers toward black and Hispanic neighborhoods (Turner and Mickelsons 1992).
13. Scholars also have been concerned about making estimates of discrimination as precise as possible to be sure that they will identify discrimination when it actually exists. As pointed out by Yinger (1986), a key step in obtaining precision is to account for unobserved factors that are shared by teammates. This step, which does not affect the estimated level of discrimination, only its statistical significance, is incorporated into most of the testing research discussed in this paper.
14. The reader may wonder why the Ondrich/Ross/Yinger upper bound estimate can exceed the simple gross estimate. The explanation, first presented in Yinger (1993), is that random events can result in the appearance of equal treatment even when the housing agent is trying to discriminate. Consider an agent who always withholds the advertised unit from blacks. If the black auditor goes first and the unit is rented before the white auditor arrives, then the audit will observe equal treatment even in this case.
15. In addition, the logit approach finds statistically significant discrimination for every type of agent behavior in these two tables except for invitations to call back for blacks and queries about tester income for Hispanics. The assumptions in the text are not required to determine statistical significance. See Ondrich, Stricker, and Yinger (1998, table 2).
16. About 1,500 tests conducted in the Chicago area in the 1990s found differences in treatment almost one-third of the time (Leachman et al. 1998), but I do not know if these tests were based on a random sample of advertisements.
17. Several of these studies also examine discrimination on the basis of familial status or handicap, protected classes that were added to the Fair Housing Act by the 1988 Amendments. See Schwemm (1992).
18. The minority and white neighborhoods apparently were not matched on very many characteristics, but the house in the minority neighborhood was selected to have a newer furnace or some other newer system. See Smith and Cloud (1997).
19. Smith and Cloud (1997, p. 103) also point out that whites who buy houses in black or Hispanic neighborhoods are affected by redlining, too. However, the NFHA study apparently did not look for this type of treatment.
20. Some scholars have implied that the Housing Discrimination Study is no longer relevant because it is based on data that were collected before these amendments were implemented. See Orlebeke (1997). I disagree with this assessment because of the testing evidence from the 1990s that is cited in this paper, but I would prefer to resolve this issue with another national testing study.
21. Although the audit or testing methodology has been refined over the years, several improvements could be made, particularly if one wants to measure the incidence of discrimination as defined by the law, not simply average differences in the treatment of two groups. For a discussion of some possible improvements, see Ondrich, Ross, and Yinger (1997).



22. Yinger (1995) provides other examples of the link between testing results and enforcement strategies.
23. Smith and Cloud (1996) argue that the best strategy would be to work for an exception in these laws for testing purposes. Some people argue that the law, which prohibits the use of false information *with intent to defraud* already contains such an exception, but no court has yet ruled on this issue.
24. For an alternative discussion of the challenges facing testing studies of loan approval, see Galster (1993).
25. I am grateful to George Galster for pointing this problem out to me.

References

- Central Alabama Fair Housing Center. 1996. *Discrimination in the Rental Housing Market: A Study of Montgomery, Alabama, 1995-6*. Montgomery, Ala.: Author.
- Fair Housing Action Center, Inc. 1996. *Greater New Orleans Rental Audit, 1996*. New Orleans: Author.
- Fair Housing Council of Fresno County. 1995. *Audit Uncovers Blatant Discrimination Against Hispanics, African Americans, and Families with Children in Fresno County*. Fresno, Calif.: Author.
- The Fair Housing Council of Greater Washington. 1997. *The Fair Housing Index*. Washington, D.C.: Author.
- Fix, Michael, George C. Galster, and Raymond J. Struyk. 1993. "An Overview of Auditing for Discrimination." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by M. Fix and R. J. Struyk, 1–49. Washington, D.C.: Urban Institute Press.
- Galster, George C. 1993. "Use of Testers in Investigating Discrimination in Mortgage Lending and Insurance." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by M. Fix and R. J. Struyk, 287–334. Washington, D.C.: Urban Institute Press.
- _____. 1990a. "Racial Discrimination in Housing Markets during the 1980s: A Review of the Audit Evidence." *Journal of Planning Education and Research* 9(3): 165–75.
- _____. 1990b. "Racial Steering by Real Estate Agents: Mechanisms and Motives." *Review of Black Political Economy* 19 (Summer): 39–63.
- Goering, John, and Ron Wienk, eds. 1996. *Mortgage Lending, Racial Discrimination, and Federal Policy*. Washington, D.C.: Urban Institute Press.
- Hakken, Jon. 1979. *Discrimination against Chicanos in the Dallas Rental Housing Market: An Experimental Extension of the Housing Market Practices Survey*. Washington, D.C.: U.S. Department of Housing and Urban Development.
- Lawton, Rachel. 1996. "Preapplication Mortgage Lending Testing Program: Lender Testing by a Local Agency." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by J. Goering and R. Wienk, 611–22. Washington, D.C.: Urban Institute Press.



- Leachman, Mike, Phil Nyden, Bill Peterman, and Darnell Coleman. 1998. *Black, White, and Shades of Brown: Fair Housing and Economic Opportunity in the Chicago Region*. Chicago: Leadership Council for Metropolitan Open Communities.
- Munnell, Alicia H., Lynn E. Browne, James McEneaney, and Geoffrey M. B. Tootell. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86 (March): 25–53.
- Newburger, Harriet. 1995. "Sources of Difference in Information Used by Black and White Housing Seekers: An Exploratory Analysis." *Urban Studies* 32 (April): 445–50.
- Ondrich, Jan, Stephen Ross, and John Yinger. 1997. "Measuring the Incidence of Discrimination Using Fair Housing Audits." Metropolitan Studies Occasional Paper No. 167, Center for Policy Research, The Maxwell School. Syracuse, N.Y.: Syracuse University [revised June 1997].
- Ondrich, Jan, Alex Stricker, and John Yinger. 1998. "Do Real Estate Brokers Choose to Discriminate? Evidence from the 1989 Housing Discrimination Study." *Southern Economic Journal*. (April): Vol. 64, 880–901.
- Orlebeke, Charles J. 1997. "Review of *Closed Doors, Opportunities Lost: The Continuing Cost of Housing Discrimination* by John Yinger." *Journal of Policy Analysis and Management*, 16 (Winter): 180–85.
- Page, Marianne. 1995. "Racial and Ethnic Discrimination in Urban Housing Markets: Evidence from a Recent Audit Study." *Journal of Urban Economics* 38 (September): 183–206.
- Ross, Stephen. 1997. "Using Testers or Auditors to Test for Mortgage Lending Discrimination." Unpublished manuscript, University of Connecticut.
- Roychoudhury, Canopy, and Allen C. Goodman. 1996. "Evidence of Racial Discrimination in Different Dimensions of Housing Search." *Real Estate Economics* 24 (Summer): 161–78.
- _____. 1992. "An Ordered Probit Model for Estimating Racial Discrimination through Fair Housing Audits." *Journal of Housing Economics* 2 (December): 358–73.
- San Antonio Fair Housing Council, Inc. 1997. *San Antonio Metropolitan Area Rental Audit, 1997*. San Antonio, Tex.: Author.
- Schwemm, Robert G. 1992. *Housing Discrimination Law and Litigation*. Deerfield, Ill: Clark, Boardman, Callaghan.
- Smith, Shanna L. and Cathy Cloud. 1997. "Documenting Discrimination by Homeowners Insurance Companies through Testing." In *Insurance Redlining: Disinvestment, Reinvestment, and the Evolving Role of Financial Institutions*, edited by G. D. Squires, 97–117. Washington, D.C.: Urban Institute Press.
- _____. 1996. "The Role of Private, Nonprofit Fair Housing Enforcement Organizations in Lending Testing." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by J. Goering and R. Wienk, 589–610. Washington, D.C.: Urban Institute Press.
- Squires, Gregory D., and William Velez. 1988. "Insurance Redlining and the Process of Discrimination." *Review of Black Political Economy* 16 (Winter): 63–75.



- Turner, Margery Austin. 1992. "Discrimination in Urban Housing Markets: Lessons from Fair Housing Audits." *Housing Policy Debate* 3(2): 185–215.
- Turner, Margery Austin, and Maris Mickelsons. 1992. "Patterns of Racial Steering in Four Metropolitan Areas." *Journal of Housing Economics* 2 (September): 199–234.
- Turner, Margery Austin, Raymond J. Struyk, and John Yinger. 1991. *Housing Discrimination Study: Synthesis*. Washington, D.C.: U.S. Department of Housing and Urban Development.
- Vartanian, Thomas P., Robert H. Ledig, and Alisa Babitz. 1995. "Disparate Impact Discrimination: Fair Lending at a Crossroads." *Consumer Finance Law Quarterly Report* (December).
- Wienk, Ronald E., Clifford E. Reid, John C. Simonson, and Frederick J. Eggers. 1979. *Measuring Discrimination in American Housing Markets: The Housing Market Practices Survey*. Washington, D.C.: U.S. Department of Housing and Urban Development.
- Wissoker, Douglas, Wendy Zimmermann, and George Galster. 1998. *Testing for Discrimination in Home Insurance*. Washington, D.C.: The Urban Institute.
- Yinger, John. 1998. "Evidence on Discrimination in Consumer Markets." *Journal of Economic Perspectives* 12 (Spring): 23–40.
- _____. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. New York: Russell Sage Foundation.
- _____. 1993. "Access Denied, Access Constrained: Results and Implications of the 1989 Housing Discrimination Study." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by M. Fix and R. J. Struyk, 69–112. Washington, D.C.: Urban Institute Press.
- _____. 1991. "Acts of Discrimination: Evidence for the 1989 Housing Discrimination Study." *Journal of Housing Economics* 1 (December): 318–46.
- _____. 1986. "Measuring Discrimination with Fair Housing Audits: Caught in the Act." *American Economic Review* 76 (December): 881–93.



Chapter 3

Adding Testing to the Nation's Portfolio of Information on Employment Discrimination

MARC BENDICK, JR.

Introduction

Beneath the surface of many current controversies about employment discrimination and its remedies lurk differences of perceptions about empirical reality: Do racial and ethnic minorities today enjoy the same job opportunities as nonminorities? How many employers deliberately treat women differently from men? Is discrimination litigation typically frivolous or well-founded?

The United States generates its answers to such questions from a portfolio of information sources. Personal experience, anecdotes, and journalism exemplify intuitive components of this portfolio; opinion polls, laboratory experiments, and statistical studies represent approaches based on more formal research. Drawing from these sources, the nation enjoys information that is reasonably accurate on some aspects of employment discrimination but seriously inaccurate on others.

Employment testing is a new source of information developed within the past decade and implemented to date only on a modest scale. This chapter

argues that testing uniquely bridges the intuitive and research components of the information portfolio. *In a world in which stories are more powerful than studies, testing generates studies that are also stories.* That characteristic gives testing unique potential to increase the effective information on which the nation bases its employment discrimination policies. Testing should be far more widely used to measure and monitor the nation's progress on this important issue.

The chapter develops this conclusion as follows. The first section summarizes the state of employment discrimination in the United States today, as portrayed by research. The next section then describes the major audiences for such information and the accuracy of their perceptions. The third section describes employment testing and its potential to close information gaps identified in the previous section. In the fourth through seventh sections, a four-part program to fulfill this potential is outlined. The final section concludes with a proposal for an annual national "report card" combining testing and non-testing data.

What Research Reveals about Employment Discrimination¹

Patterns of Unequal Employment Outcomes

One of the most well-established characteristics of the American labor market is that employment outcomes are far from equally distributed on demographic dimensions such as race, gender, age, and disability status. To illustrate this point, table 1 presents 10 indicators of labor market outcomes, ranging from unemployment rates to measures of earnings and job quality. For each indicator, the table provides, in bold type, the ratio between the indicator's value for white males and five other race/ethnicity and gender categories.

If employment outcomes in the United States were not related to workers' race/ethnicity and gender, then the bold figures would be approximately 1.0 throughout table 1. However, that is clearly not the case. For example, the unemployment rate for black males is 2.22 times that for white males; median annual earnings for Hispanic females are 56 percent of those of white males; and white females with only a high school diploma are 2.32 times as likely as corresponding white males to be employed in a service occupation.

Such differences are so well documented that their existence is not controversial. However, controversies abound concerning the explanation of these differences. Roughly, the differing positions in this debate can be divided into *employer-focused* explanations and *worker-focused* explanations.

In employer-focused explanations, the predominant cause of group differences such as those in table 1 is discrimination, conscious or unconscious, by the individuals and institutions that are the gatekeepers of employment opportunities. These gatekeepers include employers, as well as educational and training institutions, unions, job placement systems, employees' coworkers, and even the news and entertainment media that shape attitudes and perceptions. This interpretation emphasizes instances of *disparate treatment*, in which employment decisionmakers perceive, welcome, or reward persons of equal



Table 1 Selected Employment Outcomes by Race/Ethnicity and Gender, U.S. Civilian Labor Force, 1994

Employment Outcome	White Males	White Females	Black Males	Black Females	Hispanic Males	Hispanic Females
Labor Force Participation %	75.9 1.00^a	58.9 .78	69.1 .91	58.7 .77	79.2 1.04	52.9 .70
Unemployment Rate %	5.4 1.00	5.2 .96	12.0 2.22	11.0 2.04	9.4 1.74	10.7 1.98
% College Grads in Professional or Managerial Occupations	66.6 1.00	70.5 1.06	56.4 .85	68.3 1.03	— ^b —	— —
% with Only High School Diploma in a Service Occupation	8.3 1.00	19.2 2.32	19.1 2.30	32.9 3.96	— —	— —
% Represented by a Union	17.2 1.00	12.1 .70	23.3 1.35	18.1 1.05	15.5 .90	12.1 .70
% Using a Computer on the Job ^c	48.7 1.00	— —	36.2 .75	— —	29.3 .60	— —
% Allowed Flexibility in Work Schedule ^c	15.5 1.00	— —	12.1 .78	— —	10.6 .68	— —
% Covered by a Pension Plan	41.8 1.00	37.5 .90	35.6 .85	37.5 .90	24.4 .58	25.4 .61
% of Hourly Paid at or below Federal Minimum Wage ^c	6.1 1.00	— —	6.5 1.07	— —	8.6 1.41	— —
Median Annual Earnings	\$28,444 1.00	\$21,216 .75	\$20,800 .73	\$17,992 .63	\$17,836 .63	\$15,860 .56

^aFigures in bold are the ratio of the reported figure to the corresponding figure for white males.

^bDashes indicate data not available.

^cData not available by gender.

qualifications but different backgrounds differently. It also encompasses instances of *disparate impact*, in which systems and procedures that treat persons from different groups equally nevertheless result in consistently more favorable outcomes for some groups than others. To the extent that the requirement or process generating these differences is not justified by business necessity, then American law categorizes these outcomes as discriminatory as well.

The alternative worker-focused explanation typically acknowledges that instances of discrimination do occur. However, this interpretation describes such occurrences as rare and finds the principal cause of group differences in employment outcomes in the behavior of workers themselves.

In particular, this explanation focuses on differences among demographic groups in *employment qualifications*. For instance, to explain the differences in average earnings reported in the final row of table 1, this interpretation focuses on differences in educational achievement. In terms of formal educational credentials, for example, the proportion of black males who are high school graduates is only 87 percent of the corresponding proportion for white males, and the proportion who are college graduates is only 49 percent that for white males. This line of reasoning is often extended to less formally documented



aspects of employment readiness as well. For example, recent research emphasizes that, in evaluating applicants for entry-level employment, employers particularly value such “soft skills” as dependability, honesty, the ability to communicate orally and in writing, and the ability to relate to coworkers and supervisors (Holzer 1996; Murnane and Levy 1996; SCANS 1992). Proponents of worker-focused explanations often attribute the poor employment prospects of such groups as minority inner-city youth to lack of work-readiness on these dimensions.

Worker-focused explanations also emphasize differences among groups in *occupational interests*. When workers voluntarily select jobs and careers to match personal preferences, then group differences in occupational distributions might arise without employers’ discriminatory behavior. For example, according to the 1990 Census, women constitute 94.3 percent of registered nurses, but only 20.7 percent of physicians. This pattern might reflect discrimination against women by medical schools, employers, and health care consumers. But proponents of worker-focused explanations typically argue that it reflects women’s preferences as well. Specifically, they posit that women on average have a greater desire than men for jobs requiring less educational investment and imposing less work pressure, so that they can more easily pursue child-rearing (Becker 1965; Schultz and Peterson 1992; Jacobsen 1994).

Six Research-Based Generalizations

Obviously, employer-focused and worker-focused explanations both raise important points, and researchers continue to disagree concerning the quantitative balance between the two. Nevertheless, substantial consensus has been achieved among researchers on six important generalizations.

The first generalization concerns the *prevalence of discrimination* in the contemporary American labor market. In numerous studies covering a variety of racial/ethnic, gender, age, and other demographic groups, when differences in qualifications and interests are accounted for, differences in employment outcomes reduce substantially. However, in virtually no cases do they fall to zero, and in most cases not close to zero. Thus, for example, when salaries of women are statistically compared with those of men with similar education and work experience, men’s earnings typically average approximately 10 percent more than those of equally qualified women (Egan and Bendick 1994). After differences in education and experience are accounted for, racial/ethnic minorities remain underrepresented in higher-level occupations (Gill 1989). And when employees acquire additional experience, wages for younger workers increase but wages for older workers decline (Wanner and McDonald 1983). These persistent patterns make clear that, in the 1990s, discrimination continues to operate in the American labor market to a very important extent.

A second generalization concerns the *form of this continuing discrimination*. Before major federal antidiscrimination laws were enacted starting in the 1960s, it was not uncommon to encounter state and local “Jim Crow” statutes explicitly precluding racial and ethnic minorities from certain types of employment, newspaper classified advertising that divided “Help Wanted—Male” from “Help Wanted—Female,” mandatory retirement that separated older



workers from jobs they wished to retain, and workplaces in which many positions were held exclusively by members of a single race or gender group.

Despite the combined pressure of antidiscrimination laws and changing societal norms, such occurrences have by no means disappeared.² However, their prevalence has been dramatically reduced. Although it remains common to see men and women performing different jobs and receiving different pay, it is much less likely today to observe them receiving different pay for performing the same job. While it remains common to observe occupations that include women or minorities in very small numbers (“tokens”), it is much less likely today to see women and minorities entirely absent (the “inexorable zero”). And while it remains common for white males to be more likely than women or minorities to receive a job offer after being interviewed, it is much less likely for women or minorities to be refused the opportunity to be interviewed.

Reductions in blatant discrimination leave less harsh and dramatic forms of discrimination to predominate in the labor market today. These forms frequently feature, for example, multiple differences in treatment, each one of which is not crucial but whose cumulative effect places individuals on substantially different career paths (Word, Zanna, and Cooper 1974). They often derive from social relationships that limit access to information about job opportunities and applicants (Granovetter 1974; Bendick 1989b). They may reflect issues of “social comfort” and personal style that affect whose comments get listened to, who is perceived as competent, and who gets credit for accomplishments (Tannen 1994). And, as discussed below, they are inevitably rooted in assumptions about individuals based on stereotypes about that person’s demographic group. Such mechanisms may be less obvious than physically aggressive sexual harassment, racial name-calling, or posters announcing, “No Irish need apply.” However, that softer guise does not diminish their seriousness. These arrangements can powerfully distort who gets hired; what they are paid; who gets preferred assignments, training, and promotions; and who gets disciplined or dismissed (Braddock and McPartland 1987; Zwerling and Silver 1992).

The third generalization concerns the role of *stereotypes* in discriminatory behavior. Research in cognitive social psychology documents three patterns of human thought relevant to interpersonal behavior in the workplace: First, persons’ prior assumptions about group characteristics strongly influence how they perceive and judge individuals they encounter. Second, persons whose perceptions and judgments are influenced by such assumptions are often unaware of that influence and perceive themselves as unbiased. Third, the stereotypes widely held in American society are highly unfavorable toward groups traditionally discriminated against. For example, images of African Americans and Hispanics commonly held both by the general public and by employers portray them, relative to nonminorities, as less intelligent, honest, energetic, stable, and articulate and more prone to violence (Smith 1990; Neckerman and Kirschenman 1991; Culp and Dunson 1986).

The fourth generalization concerns the *information content* of employment qualifications. As noted earlier in this paper, demographic groups often differ in their possession of formal qualifications. This pattern is evident, for example, in educational attainment (years of education completed, fields of study selected,



grades awarded); work experience (length of work experience, extent of opportunities for on-the-job learning); and formal credentials (completion of organized apprenticeships, acquisition of certifications such as C.P.A.). It also often typically arises in terms of scores on paper-and-pencil tests and ratings on job interviews (Hartigan and Wigdor 1989).

But what precisely do such qualifications signify? In many cases, the relationship to employees' on-the-job performance is marginal at best. Specifically,

- Qualifications required or preferred by employers may be only weakly justified in terms of business necessity. For example, many insurance companies prefer that trainees for insurance sales have college degrees. However, they typically do not specify what enhancement in an employee's ability to sell insurance that degree is supposed to convey and have not analyzed whether persons with degrees are more successful in the sales role.
- The distinction between persons rated qualified and those not qualified may be marginal. For example, in the warehouse of a manufacturing plant in Mississippi, the company promoted workers to forklift driver from among warehouse laborers who were "qualified," meaning that they had forklift experience. Although many warehouse laborers were African Americans, the "qualified" group was all white. But that qualification could be acquired with only a single day's experience, usually gained at the company itself by being assigned to fill in for an absent regular driver.
- Research in industrial psychology concludes that most screening and rating processes routinely applied in hiring and promotional decisions have limited power to identify more promising employees. For example, personal interviews of job candidates are part of virtually every job selection process, but performance on interviews predicts only about 10 percent of the difference among hirees in subsequent job performance (Reilly and Chao 1982).

In such circumstances, differences in measured qualifications often represent less than they appear to represent.

The fifth generalization raises similar questions about *occupational interests*. As with qualifications, career aspirations often differ substantially among demographic groups. For example, a higher proportion of African Americans seek employment in the public sector than would be expected based on their representation in the overall labor force, and in opinion polls, more women than men state that they place priority on finding employment compatible with family responsibilities (Albelda 1986; Reskin 1984).

However, research indicates that such patterns of aspirations are heavily influenced by what workers perceive as realistic and often change with outreach and experience. In other words, workers' reluctance to aspire to certain occupations may not reflect strongly held personal preferences but rather the absence of demographically similar role models in that occupation, lack of exposure to the field, or reluctance by employers to make even minor, low-cost adaptations of jobs to accommodate persons with different personal preferences (Hagniere and Steinberg 1989).

The sixth and final generalization concerns the role of *well-designed personnel practices* in reducing the prevalence of discrimination. In general, discrimination is more likely in workplaces where human resource management



decisions are made informally, subjectively, “behind closed doors,” and without documentation, explicit and validated criteria, advertising of opportunities, or training for supervisors and other decisionmakers. Of course, formal rules and procedures—for example, periodic performance reviews, public postings of job vacancies, and written job descriptions—cannot by themselves guarantee the absence of bias. However, they tend to constrain extreme cases of irrationality, promote transparency of understanding between employers and employees, broaden the pool of candidates considered when opportunities arise, and help employment decisionmakers to be consistent (Cascio 1998).

Do Perceptions Match Research Findings?

With regard to employment discrimination and its remedies, important decisions are made in four principal venues. The venue of *public opinion* affects the behavior of individuals in daily interactions in the workplace, as well as their behavior as voters. The venue of *public policy* controls employment discrimination laws, the allocation of resources for their enforcement, and the range of legally permissible or legally mandated antidiscrimination initiatives. The venue of *personnel management* governs employers’ policies and decisions concerning workers’ recruitment, hiring, compensation, training, promotion, discipline, and dismissal. Finally, the venue of *litigation* adjudicates employment disputes, with rulings sometimes affecting only the parties in a specific case and sometimes affecting society more broadly.

In each of these venues, perceptions of empirical reality influence the decisions reached. These perceptions, in turn, are based on research only to the extent that research findings are effectively disseminated to the relevant audiences and that those audiences find the information credible and worthy of attention.

How effectively does empirical information concerning employment discrimination, such as was summarized in the first section, reach and influence decisionmakers in the four venues? The answer is mixed, with reasonably accurate information prevailing on some topics and misleading beliefs prevailing on others.

On three important subjects, general consensus has been achieved in American society that is consistent with the findings of research. These subjects are the extent of *past discrimination* in the American workplace, the provisions of *civil rights laws* regarding blatantly discriminatory behavior, and the incompatibility of blatantly discriminatory behavior with *current societal norms*.

This consensus is revealed in public opinion polls, which report that the majority of the American public agrees with the concept of equal opportunity in the workplace (Louis Harris 1989; Kluegel and Smith 1986). It is further demonstrated by the reluctance of either political party to question seriously the basic equal opportunity provisions of federal, state, and local antidiscrimination laws. The majority of individuals in the American workplace behave as if they understand the risks of social, managerial, or legal sanctions now associated with blatantly discriminatory behavior. Although these common understandings have not led to universal abolition of discriminatory acts, they tend to limit



them to isolated circumstances and impose self-conscious furtiveness on persons engaging in them.

Supporting this consensus, empirical information on these three points has been communicated repeatedly, extensively, and in multiple ways over several decades. With respect to race, for example, communication began in the 1960s with media coverage of dramatic incidents in the civil rights struggle, and it continues to be reinforced through symbols such as the Martin Luther King holiday and Black History Month. With respect to other groups, public service announcements urge us to hire the handicapped, workplace training warns us not to engage in sexual harassment, and equal opportunity posters are as familiar on lunchroom bulletin boards as their counterparts promoting workplace safety.

Perhaps the clearest demonstration of the consensus is provided when these norms are visibly violated. Over the past several years, employment discrimination has made front page news on several occasions. Denny's restaurants were caught engaging in discriminatory treatment of African-American customers. Physical and verbal sexual harassment surfaced at a Mitsubishi auto assembly plant. Senior executives of Texaco were tape-recorded uttering racial epithets. All these cases triggered widespread adverse publicity, threats of consumer boycotts, multimillion-dollar legal settlements, and statements of outrage from public officials. An information-based consensus had ruled these acts outside the acceptable mainstream.

No similar consensus has been achieved on three other aspects of employment discrimination.

The first is the *current prevalence* of discrimination. Public opinion polls report that, among persons who are not members of groups traditionally facing discrimination, the predominant opinion is that discrimination in employment, as in other aspects of American life, is largely a problem of the past. For example, in one nationwide sample, only 37 percent of whites thought that an African-American applicant who is as qualified as a white would be less likely to win a job that both want, and only 41 percent felt that the chances of an African American to win a supervisory or managerial position were more limited than those of counterpart whites (Louis Harris 1989). This perception contrasts sharply with the research summarized in the first section.³

The second topic on which American society has not achieved an accurate, information-based consensus is the significance of *less blatant forms of discrimination*. This topic is related to the first, for when survey respondents characterize discrimination as a problem of the past, many appear to be referring to blatant, conscious discriminatory behavior. Research, particularly that summarized under the second and third generalizations in the first section, makes clear that discriminatory acts need not be direct, dramatic, or deliberate to create major differences in employment opportunities. This insight has not been effectively communicated to many of America's voters, workers, elected officials, employers, or judges and jurors.

The final topic on which an information-based consensus is lacking concerns the role of antidiscrimination initiatives, such as affirmative action, that extend beyond ensuring equal treatment of equally qualified individuals. The research summarized in generalizations four through six in the first section



implies that employment outcomes can often be altered substantially by reexamining the qualifications required to perform jobs, exposing workers to career opportunities they might not have otherwise considered, and redesigning employment procedures and practices to make them more consistent and rational. America's voters, workers, elected officials, employers, and judges and jurors have not been made adequately aware of the role of such actions in promoting equal opportunity in the workplace. Nor have they effectively been informed that such approaches are typically the central direction of affirmative action. Instead, they have been left with misperceptions that equate affirmative action with quotas, reverse discrimination, and promotion of the unqualified (Thernstrom and Thernstrom 1997).

Testing: A New Information Source

Testing for Research Purposes

Given these information gaps, it is not surprising that public and private antidiscrimination activities have periodically come under intense attack. In particular, starting with the presidency of Ronald Reagan, public policy was marked by sharp cutbacks in the funding of antidiscrimination agencies, governmental advocacy of positions hostile to previously supported initiatives, conservative appointments to the federal bench, and Supreme Court decisions (notably, *Croson* and *Atonio*) raising the standards of proof required to support discrimination charges (Clark 1989).

Observers sympathetic to antidiscrimination and affirmative action initiatives often argued that these developments reflected a false premise that discrimination was no longer a problem in American society (Bergmann 1996). Development of employment testing is directly traceable to one observer who had the vision to see testing as a fresh response to this premise. This astute observer was James Gibson, then a senior official of The Rockefeller Foundation, who initiated an exploratory grant to the Urban Institute in 1987.

That grant underwrote development of a prototype approach to employment testing (Bendick 1989a) that drew heavily on the experience of housing testing that was then becoming well-established.⁴ This prototype was variously adapted and implemented in a series of research projects over the subsequent decade. The first studies were fielded by the Urban Institute, examining the employment experiences of Hispanics (Cross et al. 1990; Kenney and Wissoker 1994) and African Americans (Turner, Fix, and Struyk 1991). The Fair Employment Council of Greater Washington followed with studies of Hispanics (Bendick et al. 1991), African Americans (Bendick, Jackson, and Reinoso 1994), and older workers (Bendick, Jackson, and Romero 1996; Bendick, Brown, and Wall 1997). Two additional studies have been completed by researchers not involved in the initial design. Selected characteristics of these nine efforts are summarized in table 2.⁵

Table 2 also reports the key findings of these investigations. In nearly all cases,⁶ the studies document substantial discrimination in hiring in the contemporary American labor market, thereby confirming patterns described in the





Table 2 *Characteristics of Nine Testing Studies of Employment Discrimination in Hiring, 1989–96*

Characteristic	(a) African-American/White		(b) Anglo/Hispanic		(c) Older/Younger		(d) Male/Female	
	FEC ^a	UI ^a	FEC	UI	FEC	UC	FEC	Neumark
Number of Tests	149	300	498	300	79	140	102	65
Labor Market(s)	Washington	Chicago, Washington	Washington	Chicago, San Diego	nationwide	Denver	Washington	Philadelphia
Source of Job Sample	newspaper lists, walk-ins	newspaper	newspaper lists, walk-ins	newspaper	lists	newspaper	newspaper	list
Method of Application	in-person	in-person	in-person	phone, mail	mail	in-person	in-person	in-person
% Resumes	0	0	50	0	67	0	100	0
College Graduates								
Industry								
Retail	22%	23%	34%	16%	10%	37%	26%	0%
Office	5	21	46	10	34	17	21	0
Service	63	37	4	56	8	35	46	100
Other	10	19	16	18	48	11	7	0
Net Rate of Discrimination	24%	13% ^b	22%	20%	26% ^b	-10% ^c	41%	40%

^aFEC=Fair Employment Council of Greater Washington, Inc.; UI= The Urban Institute; UC= University of Colorado-Denver

^bIncludes only some modes of discrimination.

^cProtected group was treated more favorably than the non-protected group; see note 6.

first section. The final row of the table reports that, when matched pairs of job seekers with equal qualifications applied for the same job vacancy, African-American, Hispanic, older, or female applicants were treated less favorably than their white, non-Hispanic, younger, or male counterparts by a substantial proportion of employers. In the case of African-American and Hispanic job seekers, that proportion is about *25 percent*. In the case of older and female job seekers, it is about *40 percent*.

As these studies were completed, their findings were documented in scholarly books and journals. Some have also been presented in public policy forums. In particular, one Urban Institute effort (Cross et al. 1990) was sponsored by the U.S. General Accounting Office and was reported as part of that agency's congressionally mandated evaluation of the impact of the federal Immigration Reform and Control Act (IRCA). More recently, syntheses of the studies have been presented in congressional testimony, in debates surrounding California's anti-affirmative-action Proposition 209, and in the deliberations of President Clinton's Initiative on Race (e.g., Bendick 1995).

Similar syntheses have been presented on a limited number of occasions to the employer community (e.g., Bendick 1994). Some of the studies—notably those conducted by the Fair Employment Council of Greater Washington—have been formally released to the news media, which typically gave them limited coverage. The only extensive attention has arisen from testing conducted by the news media themselves, sometimes with technical assistance from organizations such as the Fair Employment Council of Greater Washington. In 1997, for example, the television news magazine *Frontline* broadcast dramatic “hidden camera” footage contrasting the experiences of a job applicant in a wheelchair with that of a nondisabled testing partner.

Testing for Litigation

Concurrently with providing support for testing-based research, The Rockefeller Foundation provided seed money to the Fair Employment Council of Greater Washington and the Washington Lawyers Committee for Civil Rights and Urban Affairs to develop testing for antidiscrimination litigation. In the early 1990s, these organizations adapted the prototype testing methodology to this purpose, conducted a series of litigation-oriented tests, and brought two suits (Boggs, Sellers, and Bendick 1993). One of these suits, filed in federal court, alleged racial discrimination in job placement by the Washington, D.C., affiliate of a nationwide employment agency, Snelling and Snelling. The other suit, filed in District of Columbia superior court, alleged sexual harassment by the proprietor of a small job placement firm. Both cases were settled with significant damages awarded to the plaintiffs, including the Fair Employment Council of Greater Washington and its testers.

Since that time, a handful of additional testing-based suits have been filed and settled by other organizations, including the Chicago Legal Assistance Foundation and the Massachusetts Commission Against Discrimination. In 1992, the federal Equal Employment Opportunity Commission (EEOC) adopted a policy of accepting discrimination charges based on tester evidence. And in 1997, both the EEOC and the other principal federal employment discrimina-



tion enforcement agency, the Office of Federal Contract Compliance Programs (OFCCP), initiated pilot projects using testing in their investigative activities. In virtually all these developments, the Fair Employment Council of Greater Washington played a role as an advocate, advisor, trainer, or contractor.

Testing's Underdeveloped Potential

Perhaps the most basic finding from a decade of employment testing is that the technique has tremendous potential to address the information gaps identified in the second section of this chapter. It can generate findings that are controlled and objective yet possess vivid persuasive power. It can document forms of discrimination that other empirical techniques cannot. It can provide unique insights into psychological and social processes and thereby lead to improved antidiscrimination practices.

Given this potential, the limited scale of testing's current use is frustrating. Throughout a decade in which issues of race and gender have been hotly contested in newspaper headlines, the voting booth, legislative bodies, the nation's highest courts, and even bloody riots, employment testing has never moved beyond an ad hoc, sporadic, hand-to-mouth scale. Only a modest body of testing-based research has been completed, and few dramatic advances in testing techniques have occurred. There has been no concerted dissemination to make testing-based insights common knowledge among the general public, public policymakers, or employers. Although some important legal precedents have been set, there has been no large volume of testing-based litigation. The conference at which this analysis was first presented itself symbolized the failure to establish a broad constituency of producers and consumers for the technique; attendance was dominated by the same small group of researchers, lawyers, and advocates who have been involved in the activity from its inception.

The moment has arrived—indeed, it is long overdue—to boost employment testing to a qualitatively different level of activity and influence. The next sections of this chapter outline four principal directions for this development.

Testing to Communicate the Current Extent of Discrimination

The first direction that should be pursued involves making public opinion and public policy more accurately informed about the *extent to which employment discrimination currently operates in the American labor market*. This understanding can support sustained or expanded antidiscrimination laws and resources for their enforcement. It can also enhance the general population's personal understanding and improve their individual behavior in the workplace.

The nine studies in table 2 provide a solid starting point for these efforts. However, the range of employment activities they encompass is far too limited. About 50 percent of the tests documented in the table were conducted in a single labor market, the Washington, D.C., metropolitan area, which cannot be assumed typical of labor markets nationwide. The range of demographic groups



whose experiences have been studied is similarly narrow. Only one study in table 2 examines gender discrimination, and that effort is limited to a single occupation, restaurant servers. No random-sample studies have been conducted on discrimination against persons with disabilities, discrimination favoring one minority group over another, or discrimination against persons from multiple protected groups (for example, women of color). In these circumstances, the research community needs to conduct additional testing studies on random samples of employers, systematically mapping local labor markets, industries, occupations, and demographic groups that have not yet been explored.

Such a series of studies would offer an opportunity to involve additional social science researchers in employment testing, thereby bringing fresh ideas to the subject as well as enhancing its visibility and credibility. Research funders might support doctoral students who wish to apply testing in their dissertations, or seek out scholars with established reputations in discrimination research but no previous experience with testing.

As additional testing studies are completed, their results need to be disseminated to the general public and public policymakers. Unlike more esoteric forms of research, testing lends itself to dramatic, visually striking, intuitively appealing presentations that can win media exposure and public attention. But to do so requires effort and public relations expertise. To maintain the credibility of testing, individual studies must be conducted in a scientifically rigorous, objective manner and published in respected scholarly outlets. However, an overall program of research and dissemination should be conceptualized as a *social marketing* initiative designed to inform public attitudes (Kotler and Roberto 1989). Research findings should find their outlet not only in scholarly journals but also in prominent news stories, visually striking television public service announcements, and congressional testimony featuring testers relating their individual experiences.

Testing to Reveal the Subtleties of Contemporary Discrimination

A second direction for testing should be to provide more accurate information to the general public and public policymakers concerning the *prevalence and significance of less blatant forms of discrimination*. As was discussed in the first section, although such discrimination often operates indirectly and without intent, it is powerfully discriminatory nonetheless.

Not only information on the prevalence of discrimination but also improved understanding of discrimination's subtler forms can improve the interpersonal behavior of individuals in the workplace. At a public policy level, it is also likely to sustain antidiscrimination laws and promote resources for their enforcement. In particular, for reasons discussed in the second section, it is likely to enhance public understanding of, and support for, actions that go beyond equal opportunity, such as affirmative action.

Testing can be particularly useful in examining subtle and complex forms of discrimination because of the unique detail it provides on psychological



processes and interpersonal interactions in the workplace. However, the field methods and analytical procedures implemented in testing studies to date have been too primitive to exploit this potential fully. Advanced methods for recording data, including hidden tape recorders and cameras, have been shown to be feasible but have not been systematically applied. More sophisticated procedures for analyzing testing experiences could be drawn from state-of-the-art concepts in linguistics and cognitive social psychology, but these have been attempted only on a preliminary level (Bendick 1996). An agenda for employment testing must include upgrading testing methodology to take advantage of these underutilized opportunities.

Testing to Improve Employers' Personnel Management Practices

A third direction in development involves *repackaging testing findings to enhance their use by employers*. Employers directly control much of what occurs in the workplace through their human resource management policies and through selection and training of the managers who implement these policies. Litigation represents one strategy for focusing employer attention on employment discrimination. Providing employers with information on problems in their workforce and opportunities to improve efficiency and profitability represents an important alternative approach. Testing has substantial underutilized potential to support the latter strategy.

In the 1990s, antidiscrimination efforts in the workplace voluntarily initiated by employers are often labeled *managing diversity*. With their goal of enhancing the productivity of employers' increasingly diverse workforces, these activities are often largely separate from traditional equal employment opportunity (EEO) and affirmative action programs designed to comply with government requirements (Thomas 1991; Jackson et al. 1992; Bendick, Egan, and Lofhjelm 1998). In initiatives to manage diversity, information such as that which can be generated through testing can play both a *motivating* role and a *facilitating* one. In the former role, information that makes higher-level executives aware of problems of discrimination and its adverse effects on employees can increase the likelihood that firms will invest in such activities; in the latter role, information supplied to diversity management trainers, organizational development consultants, and other staff implementing these initiatives can increase the effectiveness of their efforts.

To some extent, employers would absorb testing-based information from dissemination efforts targeting the general public such as were discussed in the fourth and fifth sections. However, the importance of this audience justifies more targeted outreach, including the following three initiatives.

First, testing results need to be distributed through information channels to which employers pay particular attention. Many employers' most important information source is the trade press covering their own industry. Many executives follow *Progressive Grocer* or *Iron Age* with greater intensity than they devote to general news media or even the generic business press such as the



Wall Street Journal. Although extra effort is required to write articles or make conference presentations for a series of narrow audiences rather than one broad audience, such efforts may be necessary to communicate in ways that are relevant to the target audience. Furthermore, to prepare for such eventual dissemination, testing studies might target specific industries, such as banks, employment agencies, or construction firms.

Second, the findings of testing research need to be translated into formats—such as training handouts, videotape presentations, and learning exercises—for daily use by diversity trainers and organization development consultants. To develop such products, testing researchers need to team with specialists in the development of such materials. This audience seldom reads materials published by the Urban Institute Press, however worthy, but they routinely purchase training materials from the American Management Association or the Society for Human Resource Management.

Third, testing practitioners might assist employers to implement testing within their own organizations. Many employers, especially large ones, routinely conduct in-house surveys or focus groups to measure employee satisfaction and identify workplace problems. These firms also often use testing-like approaches, such as “secret shoppers,” to monitor customer service. With technical assistance from researchers familiar with testing methodology, some imaginative employers might add testing to their sources of information about their own workplaces.

Testing to Strengthen Antidiscrimination Litigation

The fourth and final direction for development of testing’s potential involves *testing to support employment discrimination litigation*. Although such tests must be conducted with the same objectivity and care as testing for research, it is often appropriate to adapt their design to the requirements of the legal process. For example, rather than being applied to a random sample of employers, litigation-oriented tests may target firms suspected of discrimination, and one firm may be tested repeatedly to document its behavior thoroughly (Boggs, Sellers, and Bendick 1993).

Testing is generally not feasible for *posthiring* forms of discrimination, such as those involving employee assignments, compensation, promotions, or terminations; these aspects of employment involve decisions about persons already known to employers. However, testing is well suited to examining employers’ *hiring* practices. This match is fortunate because hiring discrimination is often difficult to document without testing. A job applicant who is told that a vacancy has already been filled or has been filled by someone more qualified seldom has adequate information to challenge these statements. Currently, claims of hiring discrimination constitute only about 6 percent of complaints lodged annually with the EEOC (Bendick, Jackson, and Reinoso 1994). Testing can provide more thorough monitoring of this important aspect of employment.

To implement litigation-oriented hiring testing on a wide scale will require development of employment testing capabilities in multiple local antidiscrimination organizations.⁷ In particular, nonprofit fair housing councils operate in



many locales, and many have experience testing for housing discrimination. A campaign to expand their agendas to employment discrimination could be pursued, offering these organizations training, technical assistance, and perhaps modest startup resources. Such efforts have been pursued by the Fair Employment Council of Greater Washington to a modest degree, such as organizing one nationwide training conference (Fair Employment Council 1993). However, only a far more sustained and deliberate effort is likely to achieve substantial results.

Public antidiscrimination employment agencies—notably the EEOC, OFCCP, and their state and local counterparts—can also implement testing as part of their investigative processes. As was discussed in the third section, one state agency, the Massachusetts Commission Against Discrimination, has done so on several occasions, and the EEOC and OFCCP are currently conducting pilot projects. These steps move in the right direction, but painfully slowly.

As part of their routine activities, public antidiscrimination agencies such as the EEOC have access to important data on the employment practices of individual employers. The agencies receive worker complaints alleging discriminatory behavior. They also receive periodic reports (such as the well-known EEO-1 forms) on which firms report the representation of different demographic groups among their employees. These agencies commonly use such data internally, with varying degrees of sophistication in their analyses, to identify potential targets for investigation. Following this same approach, investigations involving testing by these agencies can be targeted the same way. That procedure would raise the probability that testing will be efficiently targeted on discriminatory firms. In addition, it would prepare for litigation in which testing evidence and nontesting evidence are both presented.

Public agencies could substantially enhance private testing-based enforcement efforts if they were to make public some of the same information currently used internally. For example, OFCCP could publish data on the demographic characteristics of the workforce at individual firms that are government contractors. Or the EEOC could provide tabulations of the number of discrimination complaints lodged against individual firms. Strategically minded private enforcement agencies could use such information to target their testing efforts for maximum effect.

Combining testing and nontesting information represents one way to incorporate testing into a broader *strategy for EEO enforcement*. It is not the only way, however. In employment discrimination litigation, as in litigation in general, one necessary ingredient is a plaintiff who has suffered injury and has standing to sue. Employment discrimination enforcement is often hamstrung by mismatches between the availability of plaintiffs and the seriousness of employers' discriminatory behavior. Public agencies, such as the EEOC, labor under backlogs of tens of thousands of cases, that, although meritorious, affect only one or a few individuals. Concurrently, these agencies, nonprofit antidiscrimination organizations, or private attorneys may be aware of egregious cases of *systemic* discrimination affecting hundred or thousands of workers but cannot pursue these cases for lack of appropriate plaintiffs. In testing, the testing organization and testers who experienced discrimination during their tests can fill the role of plaintiffs. In that circumstance, testing permits public and



private enforcement resources to be targeted toward cases where discrimination problems are the most serious rather than those where plaintiffs are the most vocal.

Conclusion: Testing as the Core of an Annual National Report Card

Although the four approaches described in the previous sections would attack employment discrimination in important ways, even the four together may fail to gain for the issue the high visibility and sustained attention it requires and deserves. A fifth, capstone initiative is needed to ensure a prominent place on the national agenda. For this role, I propose an annual “national report card.” By this phrase, I mean a research report to be released at the same time every year with extensive press coverage,⁸ setting forth quantitative indicators of the current state of employment discrimination in the nation.

Some of the indicators in the report card should be generated through testing—in particular, the proportion of tests conducted that year on a random sample of employers nationwide⁹ in which employers were observed behaving in a discriminatory manner. Examples of these indicators are presented in the final row of table 2.

However, because testing can usually be applied only to hiring, and even then most easily only to entry-level jobs, this technique can provide only part of the data that should be reported. The report card should also incorporate information from at least five nontesting sources, as illustrated in table 1.

- *Earnings* of workers with different demographic backgrounds (for example, comparisons between the median annual earnings of white males and those of other gender and racial/ethnic groups). These figures are already generated and published annually by the U.S. Bureau of Labor Statistics from its monthly Current Population Survey.
- *Unemployment rates* for different demographic groups (as well as related measures of labor force participation, such as employment-to-population ratios). As with earnings, these data are already available from the U.S. Bureau of Labor Statistics.
- *Employment representation*, such as the proportion of women and racial/ethnic minorities employed in different occupations. These figures can be computed from data collected annually by the EEOC from all large employers and government contractors (e.g., U.S. EEOC 1997).
- *Acquisition of employment credentials* (for example, the numbers of women and minorities receiving degrees in fields in which they have been traditionally underrepresented and the proportion of women and minorities among persons acquiring work-related credentials such as C.P.A. or journeyman status in the construction crafts). Suitable data could be acquired from federal agencies (such as the U.S. Department of Education), state licensing boards, or trade and professional associations.



- *Discrimination disputes*, such as the number of complaints filed annually with the federal EEOC and its state and local counterpart agencies (already tabulated by the EEOC); or the number of discrimination lawsuits filed in federal courts (already tabulated by the Administrative Office of the Federal Courts).

The impact of this annual study would be greatest if the report card had three characteristics. First, each indicator should be comparable from year to year, so that progress over time (or lack of it) can be readily observed. Second, the report should monitor the experiences of all major groups traditionally subject to discrimination in the workplace—not only racial and ethnic minorities but also women, older workers, and persons with disabilities. Third, this employment report should be part of a broader system generating parallel reports on discrimination in other aspects of daily life, including housing, education, retail sales, financial services, and public accommodations.

Such an annual report would be an appropriate capstone for the testing approach to employment discrimination described throughout this paper. Like testing, it can be broad in scope but grounded in facts, rigorous in method but vivid in presentation, and credible to researchers but relevant to advocates. Like testing, it would represent an important addition to the nation’s portfolio of information on employment discrimination.

Endnotes

1. This section is based in part on Bendick (1997). For some of the vast literature underlying this discussion, see Ehrenberg and Smith (1997, ch. 12) and Bendick (1996). Table 1 is based on U.S. Bureau of the Census (1995).
2. Their continued presence is documented by the continuing flow of antidiscrimination litigation that is won by plaintiffs or settled with substantial damages (e.g., Watkins 1997), by the continuing flow of complaints lodged annually with the federal Equal Employment Opportunity Commission and its state and local counterpart agencies, by research based on personal experiences of discrimination (Feagin and Sikes 1994), and by statistical studies. That final category is exemplified by a survey of newspaper classified advertising, which found that nine percent of job vacancy announcements contained discriminatory wording, such as specifying the age or gender of desired applicants (Kohl 1990).
3. This perception also clashes with the perceptions of the groups traditionally facing discrimination, who predominantly characterize discrimination as an ongoing problem. In the poll cited in the text, for example, more than 80 percent of African-American respondents agreed with the first statement, and 62 percent agreed with the second.
4. Prior to this date, only a handful of very preliminary studies had applied testing to employment (Culp and Dunson 1986; Riach and Rich, 1991–92).
5. Table 2 is based on the sources cited in this paragraph.
6. The only research that failed to find substantial, statistically significant discrimination is that of James and DelCastillo (1992). However, this work has been heavily criticized for methodological flaws (Fix and Struyk 1993, pp. 407–13) and has never been accepted for publication in a refereed journal.
7. It will also require cultivation of favorable case law, a process that has begun with the two cases brought by the Fair Employment Council of Greater Washington (Boggs, Sellers, and Bendick 1993). Toward this end, strategic coordination needs to be maintained among litigators applying testing, particularly in the earliest cases.



8. For example, it might be released at congressional hearings at which cabinet-level federal officials would testify, perhaps on the model of extensively reported appearances by the chairman of the Federal Reserve Board before the Congressional Joint Economic Committee.
9. Although generating national rates of discrimination that are comparable from year to year would be one important objective of testing, another goal might be generation of rates for individual metropolitan areas. To support both goals within a reasonable budget, a sampling strategy might be used in which testing is conducted in a different subset of metropolitan areas each year, with each area tested periodically (for example, once every five years).

References

- Albelda, R. 1986. "Occupational Segregation by Race and Gender, 1958–1981." *Industrial and Labor Relations Review* 39 (April): 404–11.
- Becker, Gary. 1965. "A Theory of the Allocation of Time." *Economic Journal* 75: 493–517.
- Bendick, Marc, Jr. 1997. *Access, Diversity, and Civil Rights Issues in the Development of Skills Standards*. Washington, D.C.: Bendick and Egan Economic Consultants, Inc., for the U.S. National Skills Standards Board.
- _____. 1996. *Employment Practices and Employment Discrimination: A Bibliography Combining Economic, Managerial, and Behavioral Science Research*. Washington, D.C.: Fair Employment Council of Greater Washington, Inc.
- _____. 1995. "Research Evidence on Racial/Ethnic Discrimination and Affirmative Action in Employment." Testimony before the Assembly Committee on the Judiciary, California State Legislature, May 4, 1995.
- _____. 1994. "Testing for Age Discrimination in Hiring: Lessons for Employers." Presentation to the National Conference of the Society for Human Resource Management, June 14, 1994.
- _____. 1989a. *Auditing Race Discrimination in Hiring: A Research Design*. Washington, D.C.: Bendick and Egan Economic Consultants, Inc., for the Urban Institute.
- _____. 1989b. "Matching Workers and Job Opportunities." In *Rethinking Employment Policy*, edited by D. Bawden and F. Skidmore, 81–108. Washington, D.C.: Urban Institute Press.
- Bendick, Marc, Jr., Lauren E. Brown, and Kennington Wall. 1997. *No Foot in the Door: An Experimental Study of Employment Discrimination against Older Workers*. Washington, D.C.: Fair Employment Council of Greater Washington, Inc.
- Bendick, Marc, Jr., Mary Lou Egan, and Suzanne Lofhjelm. 1998. *The Documentation and Evaluation of Antidiscrimination Training in the United States*. Geneva: International Labour Office.
- Bendick, Marc, Jr., Charles Jackson, and Victor Reinoso. 1994. "Measuring Employment Discrimination through Controlled Experiments." *Review of Black Political Economy* 23: 25–48.
- Bendick, Marc, Jr., Charles Jackson, and J. Horacio Romero. 1996. "Employment Discrimination against Older Workers: An Experimental Study of Hiring Practices." *Journal of Aging and Social Policy* 8: 25–46.



- Bendick, Marc, Jr., Charles Jackson, Victor Reinoso, and Laura Hodges. 1991. "Discrimination against Latino Job Applicants: A Controlled Experiment." *Human Resource Management* 30: 469–84.
- Bergmann, Barbara. 1996. *In Defense of Affirmative Action*. New York: Basic Books.
- Boggs, Roderick, Joseph Sellers, and Marc Bendick, Jr. 1993. "Use of Testing in Civil Rights Enforcement." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael Fix and Raymond Struyk, 345–76. Washington, D.C.: Urban Institute Press.
- Braddock, Jomills, and James M. McPartland. 1987. "How Minorities Continue to Be Excluded from Equal Employment Opportunities: Research on Labor Market and Institutional Barriers." *Journal of Social Issues* 43: 5–39.
- Cascio, Wayne. 1998. *Applied Psychology in Human Resource Management*. Upper Saddle River, N.J.: Prentice Hall.
- Clark, Leroy. 1989. "Ensuring Equal Opportunity in Employment through Law." In *Rethinking Employment Policy*, edited by D. Bawden and F. Skidmore, 173–204. Washington, D.C.: Urban Institute Press.
- Cross, Harry, et al. 1990. *Employer Hiring Practices: Differential Treatment of Hispanic and Anglo Job Seekers*. Washington, D.C.: Urban Institute Press.
- Culp, Jerome, and Bruce Dunson. 1986. "Brothers of a Different Color: A Preliminary Look at Employer Treatment of Black and White Youth." In *The Black Youth Employment Crisis*, edited by Richard Freeman and Harry Holzer, 233–59. Chicago: University of Chicago Press.
- Egan, Mary Lou, and Marc Bendick, Jr. 1994. "International Business Careers in the United States: Salaries, Advancement, and Male-Female Differences." *International Journal of Human Resource Management* 5: 33–50.
- Ehrenberg, Ronald, and Robert Smith. 1997. *Modern Labor Economics*. Reading, Mass.: Addison Wesley Longman.
- Fair Employment Council. 1993. *EEO Testing Manual*. Washington, D.C.: Fair Employment Council of Greater Washington.
- Feagin, Joe R., and Melvin P. Sikes. 1994. *Living with Racism: The Black Middle Class Experience*. Boston: Beacon Press.
- Fix, Michael, and Raymond Struyk, eds. 1993. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: Urban Institute Press.
- Gill, Andrew. 1989. "The Role of Discrimination in Determining Occupational Structure." *Industrial and Labor Relations Review* 42: 610–23.
- Granovetter, Mark. 1974. *Getting a Job: A Study of Networks and Careers*. Cambridge, Mass.: Harvard University Press.
- Hagniere, L., and R. Steinberg. 1989. "Nontraditional Training for Women: Effective Programs, Structural Barriers, and Political Hurdles." In *Job Training for Women, The Promise and Limits of Public Policy*, edited by S. Harlan and R. Steinberg, 333–58. Philadelphia: Temple University Press.



- Hartigan, John A., and Alexandra Wigdor, eds. 1989. *Fairness in Employment Testing, Validity Generalization, Minority Issues, and the General Aptitude Test Battery*. Washington, D.C.: National Academy Press.
- Holzer, Harry. 1996. *What Employers Want: Job Prospects for Less-Educated Workers*. New York: Russell Sage Foundation.
- Jackson, Susan, et al. 1992. *Diversity in the Workplace*. New York: Guilford Press.
- Jacobsen, Joyce. 1994. *The Economics of Gender*. Cambridge, Mass.: Basil Blackwell.
- James, Franklin, and Steve DelCastillo. 1992. *We May Be Making Progress toward Equal Access to Jobs: Evidence from Recent Audits*. Unpublished manuscript, University of Colorado at Denver.
- Kenney, Genevieve, and Douglas Wissoker. 1994. "An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers." *American Economic Review* 84: 674–83.
- Kluegel, J. R., and E. R. Smith. 1986. *Beliefs about Equality: Americans' Views of What Is and What Ought to Be*. Hawthorne, N.Y.: Aldine de Gruyter.
- Kohl, J. 1990. "Wanted: Recruitment Advertising That Doesn't Discriminate." *Personnel* 68: 18–26.
- Kotler, Philip, and Eduardo Roberto. 1989. *Social Marketing: Strategies for Changing Public Attitudes*. New York: The Free Press.
- Louis Harris. 1989. *The Unfinished Agenda on Race in America*. New York: NAACP Legal Defense and Education Fund.
- Murnane, Richard, and Frank Levy. 1996. *Teaching the New Basic Skills*. New York: The Free Press.
- Neckerman, Joleen, and Kathryn Kirschenman. 1991. "'We'd Love to Hire Them, but. . .': The Meaning of Race for Employers." In *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson, 203–34. Washington, D.C.: Brookings Institution.
- Neumark, David. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics* XX: 915–40.
- Reilly, R., and G. Chao. 1982. "Validity and Fairness of Some Alternative Employee Selection Procedures." *Personnel Psychology* 35: 1–62.
- Reskin, Barbara, ed. 1984. *Sex Segregation in the Workplace: Trends, Explanations, Remedies*. Washington, D.C.: National Academy Press.
- Riach, P. A., and J. Rich. 1991–92. "Measuring Discrimination by Direct Experimental Methods: Seeking Gunsmoke." *Journal of Post-Keynesian Economics* 14: 143–50.
- SCANS [Secretary's Commission on Achieving Necessary Skills]. 1992. *Learning a Living: A Blueprint for High Performance*. Washington, D.C.: U.S. Department of Labor.
- Schultz, V., and S. Peterson. 1992. "Race, Gender, Work, and Choice: An Empirical Study of the Lack of Interest Defense in Title VII Cases Challenging Job Segregation." *University of Chicago Law Review* 59: 1073–81.



- Smith, Tom W. 1990. *Ethnic Images*. Chicago: National Opinion Research Center.
- Tannen, Deborah. 1994. *Talking 9 to 5*. New York: William Morrow and Co.
- Thernstrom, Stephan, and Abigail Thernstrom. 1997. *America in Black and White, One Nation, Indivisible*. New York: Simon and Schuster.
- Thomas, R. Roosevelt, Jr. 1991. *Beyond Race and Gender*. New York: American Management Association.
- Turner, Margery, Michael Fix, and Raymond Struyk. 1991. *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring*. Washington, D.C.: Urban Institute Press.
- U.S. Bureau of the Census. 1995. *Statistical Abstract of United States, 1995*. Washington, D.C.: U.S. Government Printing Office.
- U.S. EEOC. 1997. *Job Patterns for Minorities and Women in Private Industry 1996*. Washington, D.C.: U.S. Government Printing Office.
- Wanner, R., and L. McDonald. 1983. "Ageism in the Labor Market: Estimating Earnings Discrimination against Older Workers." *Journal of Gerontology* 38: 738–44.
- Watkins, Steve. 1997. *The Black O: Racism and Redemption in an American Corporate Empire*. Athens, Ga.: University of Georgia Press.
- Word, C. O., M. P. Zanna, and J. Cooper. 1974. "The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction." *Journal of Experimental Social Psychology* 10: 109–20.
- Zwerling, Craig, and Hilary Silver. 1992. "Race and Job Dismissal in a Federal Bureaucracy." *American Sociological Review* 57: 651–60.



Chapter 4

Racial Discrimination in “Everyday” Commercial Transactions: What Do We Know, What Do We Need to Know, and How Can We Find Out?

PETER SIEGELMAN

Introduction

In 1992, ABC TV took two young testers, one black and one white, and covertly filmed what happened to them as they spent a week (pretending to be) going about the business of everyday life—looking for a job, negotiating to buy a car, going shopping, trying to find an apartment, and so on (*Prime Time Live* 1992). The 20-minute segment vividly documented many instances of discriminatory treatment: the black tester was ignored in a shoe store, while the white tester received instant and friendly service; the black tester was followed as a suspected shoplifter in a record store, while his partner was accorded normal, courteous treatment; the black tester was quoted a much higher price than his partner for the identical car.

But while the documentary provided incontrovertible evidence of the existence of discriminatory treatment in many kinds of commercial transactions, it did not address the important question of how common such discrimination

is. We still do not have a good answer to this question. Although there is now a large body of research on the frequency and amount of discrimination in what are arguably the two most important markets in which most of us participate—employment and housing—we know very little about discrimination in other kinds of transactions.¹

It is hard to claim that discrimination in restaurants or shopping is as important as discrimination in employment or housing. The latter two activities are central to a person's life chances in ways that the former two clearly are not. But it also seems clear that discrimination in everyday transactions imposes significant psychological costs on its victims and is a clear violation of our civil rights laws. There is thus ample justification both for wanting to know more about it and for doing something to prevent it.

This paper has two goals: to summarize what little we know about racial discrimination in everyday commercial transactions (which I will loosely refer to as “public accommodations”) such as buying a car, hailing a taxicab, or eating a restaurant meal; and to make some practical suggestions for how we should go about testing for discrimination in these markets.² I consider questions of technique, but also discuss some broader issues such as the justifications for testing and some of the policy and enforcement implications of conducting testing in these areas.

The second section discusses discrimination in the market for new cars, one of the few areas in which we actually have data from an audit study. As the study (of which I was a co-author) has been subject to criticism for being “unrealistic,” I devote some attention to defending its results. The third section briefly discusses another audit study that measured discrimination in the market for taxicab service in Washington, D.C. The fourth section considers discrimination in public accommodations more broadly (restaurants, shopping, etc). There have been no systematic studies of this topic, so one can rely—cautiously—only on survey evidence, isolated journalistic narratives, and judicial opinions in public accommodations cases.

Perhaps the most important conclusion of this paper is that discrimination is likely to differ in form, motive, intensity, and effects across the various markets that comprise the category “public accommodations” or “everyday commercial transactions.” No single theory is likely to explain the wide variety of observed discrimination(s), and no single method is appropriate for studying such a heterogenous phenomenon.

Table 1 sketches a crude taxonomy that may be useful in analyzing the prevalence and severity of discrimination. It classifies markets along two dimensions: whether discrimination is “visible” or not, and whether discrimination takes the form of higher prices as opposed to denial or degradation of some service or opportunity.

Cell 1 of the table consists of those markets in which (a) discrimination takes the form of higher prices charged to minorities, and (b) an individual buyer finds it difficult or impossible to ascertain whether the price paid was high or low relative to other buyers. Interbuyer price comparisons are ruled out because the good or service being sold is essentially heterogenous across buyers (e.g., cars of different models, or with different options), or because consumers bargain individually with sellers over the price or other important



Table 1 *Classification of Markets by Type and Visibility of Discrimination*

	Discrimination Is Hidden	Discrimination Is at Least Partially Overt
Discrimination Takes the Form of Higher Prices	1 A. New Car Sales B. TV Repair?; Appliance Sales?; Home Repair?; Auto Repair?	2 None?
Discrimination Takes the Form of Denial or Degradation of Service or Opportunity	3 Housing; Employment	4 Restaurants; Car Rentals; Shopping; Taxicabs?

aspects of the sale. The best-studied example is the market for new cars, but cell 1.B suggests that there are other markets with similar structural characteristics, such as television or home repairs.

Because it is virtually impossible for an individual consumer to detect discrimination in such markets, and because higher markups presumably raise sellers’ profits, discrimination is likely to be relatively more frequent and severe in the markets described by cell 1. And indeed, there is substantial evidence—from an audit study, from a conventional statistical analysis of transactions data, and from pending litigation—of racial price discrimination in the sale of new cars: blacks are asked, on average, to pay anywhere from \$400 to \$1000 more than white males for identical vehicles.

Cell 2 describes markets in which there is open or overt discrimination on the basis of price. It is blank because—as far as I know—there are no examples of posted prices that are higher for blacks than for whites.³ Cell 3 contains markets in which discrimination takes the form of a covert denial of access to an opportunity or service. Preeminent examples are housing and employment, where victims who are not offered a job or not shown an apartment because of their race are usually unaware of what has happened to them, or why. As discrimination in these markets has generated substantial literatures that are discussed by other papers at this conference, I will ignore them here.

Finally, cell 4 encompasses markets where discrimination is overt and takes the form of a denial or degradation of service. It includes restaurants that refuse to seat black customers, stores that harass black shoppers, taxicabs that refuse to pick up black customers, and so on.⁴ We know very little about the prevalence or severity of discrimination in these markets, but it seems plausible a priori that it is less widespread and less serious than in cell 1. Because the type of discrimination in cell 4 is easier to detect, customers are in a better position to use legal or market remedies when they encounter it. And discrimination that results in the denial of service to patrons willing and able to pay for it will often decrease sellers’ profits, rather than presumptively increasing profits as is the case in cell 1.



These predictions are borne out by the meager available evidence. In several kinds of everyday transactions—such as eating at restaurants and shopping—the probability of discrimination *per unit of exposure* (restaurant visit, shopping trip) appears from survey evidence to be roughly 1 to 5 percent, which is substantially lower than in cars, taxicab service, employment, or housing. Despite the relatively low probability of encountering discrimination on any given shopping trip or restaurant visit, the frequency with which any *individual* experiences discriminatory treatment is relatively high: 10 to 30 percent of blacks report one or more discriminatory experience in a given month. This apparent paradox results from the fact that individuals do a lot of shopping and restaurant dining, so even though discrimination is relatively unlikely on any given trip, it is almost certain to occur if enough trips are taken.

Bargaining for a New Car

I start by reviewing the evidence obtained from an audit study of race and gender discrimination in new car negotiations conducted by Ian Ayres and myself in 1990–91.⁵ The data reveal that dealers quoted significantly lower prices to white males than to black and/or female test buyers, even though the testers closely resembled each other in dress and general demeanor and followed an identical bargaining script.

A possible weakness of testing in this context is that it was impossible to actually purchase the cars for which the testers were negotiating. Hence, some critics have argued that our results do not accurately reflect the amount of discrimination found in bargaining that culminated in genuine purchases. However, subsequent evidence from a study of actual purchases, and from pending litigation, appears to broadly confirm the conclusions we reached.⁶

Audit Results

The techniques used in auditing car dealerships were described in our earlier articles, and I will not discuss them in detail here. We relied on the standard procedures of selecting testers with similar observable attributes (age, education, “appearance”) and further homogenizing by giving them false biographies and training them to follow a uniform bargaining script. Varying pairs of testers,⁷ one of whom was always a white male, were then sent to negotiate for new cars at randomly selected Chicago-area dealerships. Testers were not told of the true purpose of the study, and did not know that more than one of them would visit each dealership.

Column 1 of table 2 summarizes the study’s key results. In brief, white male testers were able to negotiate an average final markup of roughly \$560, while white females were quoted a final price that was roughly \$130 higher than this (controlling for unobservable, dealership-specific effects). Although this disparity was not statistically significant, the black testers in our study negotiated final offers that were much higher than their white male counterparts. Black female testers were asked to pay an additional \$400, and black males an addi-



Table 2 *Estimated Price Premium over White Males in Two Studies of Markups on New Cars, by Demographic Group*

Demographic Group	Audit Data (Ayres/Siegelman)	Survey Data (Goldberg)	Standardized Difference ^a
White Females	129 (123) [53]	129 (117) [259?]	0.0
“Minority” Females	405 ^b (116) [60]	426 (525) [14?]	0.04
“Minority” Males	\$1,060 ^b (142) [40]	\$274 (263) [52?]	2.6 ^b

Notes and Sources:

Column 1, Ayres and Siegelman (1995, table 2, col. 4).

Column 2, Goldberg (1996, table 2, col. 1). Goldberg does not give separate Ns for race/gender subgroups; these were inferred from the numbers of men, women, minorities, and whites given in table 4. Goldberg’s “Minority” categories include all nonwhites, while Ayres/Siegelman used blacks only. Ayres/Siegelman controlled for dealership-specific errors with a fixed effects specification, and for other sources of heterogeneity through the audit design. Goldberg controlled for a large variety of model and buyer characteristics.

Standard errors in parentheses. Number of observations in brackets.

^aDifference in markups divided by its standard error, $\sqrt{\sigma_a^2 + \sigma_g^2}$, where σ_i is the standard error from study *i*.

^b= significantly different from zero at the 5% level.

tional \$1,060 over what white males were quoted for similar cars at the identical dealerships. (These figures represent markups of 1 to 9 percentage points over what white males were asked to pay.) The black male and black female results are economically meaningful and statistically significant, and all the results are robust—alternative specifications and different statistical tests do not alter the basic findings.

How Realistic Are the Audit Findings?

Evidence from “Nearly Completed” Transactions

Only about 20 percent of the tests ended with a seller attempting to accept a tester’s offer; the remainder concluded when the parties reached a bargaining impasse and hence may not reflect what occurs in actual sales, which of course are all consummated bargains.⁸ For obvious reasons, looking at actual transactions is impossible with audit data in this context. But we can focus on those tests in which the seller tried to accept a tester’s offer, which are as close to completed transactions as it is possible to get with test buyers. Using only these attempted acceptances, we found essentially the same pattern of discriminatory markups as when all tests were used. Moreover, dealers were more likely to attempt to accept offers made by white males than by the other tester types: This means that our estimate of the discriminatory premiums are understated, as sellers might have been willing to make even lower offers to white males.

In sum, no internal evidence suggests that the lack of actual transaction data caused us to overstate the discriminatory premium.⁹



Evidence from Survey Data on Transaction Prices

Additional evidence about the pattern of markups in actual transactions comes from a recent regression study by Goldberg (1996), who used data from the Consumer Expenditure Survey, a nationwide representative sample of households that is used to compute the Consumer Price Index. Although *derived* from actual transactions, the CES data used by Goldberg are not true “transactions prices,” but are based on consumers’ recollections of what they paid for their car. Thus, although the CES data do offer an important perspective on the operation of the market for new cars, they are subject to several important limitations, discussed at length below.

Even though Goldberg (1996, 624) characterized her results as “quite different from the ones reported by . . . Ayres and Siegelman,” columns 2 and 3 of table 2 demonstrate some striking similarities between the two sets of findings. Goldberg’s estimates of the discriminatory premiums paid by white females and “minority” females are virtually identical to ours.¹⁰ The only difference between columns 1 and 2 is that Goldberg found black males paying a much smaller premium than we did, and none of her results are statistically significant, whereas ours were, at least for the black testers. Because there are at least six dimensions on which our audit data allowed for more precise measurement and better controls than the survey Goldberg used, her failure to obtain statistically significant results is not surprising and should not be taken as evidence against the existence of discrimination in new car sales.¹¹

Goldberg’s estimated black male premium *is* at odds with our audit findings, as it is only about one-fourth the size of the discriminatory premium we found (although at \$275, her estimate is still economically meaningful). Although various technical explanations might reconcile Goldberg’s estimated black male premium with our substantially higher estimate,¹² none of them seems sufficient to explain the \$800 difference between Goldberg’s estimate and ours, which remains something of a puzzle.

One possible explanation is that even though black customers face higher prices at most dealerships (as demonstrated in the audit results), they do not in the end pay substantially higher prices, as Goldberg’s black male result suggests. This could happen if black male shoppers who face discrimination at *some* dealerships “solve” this problem by searching longer and harder for those less-discriminatory dealerships that will offer them a better deal. By doing so, they may succeed in offsetting some of the discriminatory premium detected in audit studies, but only by paying higher search costs than white consumers.¹³

Looking only at prices paid by minorities in actual transactions thus ignores the other margins on which the effects of discrimination can be felt. Goldberg (1996, 643) mentions this point, only to dismiss it because the black testers in our study did not receive better offers at dealerships in black neighborhoods. But just because black testers didn’t receive better offers in black neighborhoods does not imply that black customers would not find it in their interest to search for nondiscriminatory dealers; quite the contrary.¹⁴

Finally, regardless of which estimate of the black male premium turns out to be more accurate, it is important to remember that at \$275, Goldberg’s estimate is still economically significant. While the two disparate estimates of the black male premium poses an open question for further research, both our audit study and Goldberg’s analysis of survey data show that discrimination in the market



for new cars is a reality. In my view, Goldberg's study should be interpreted as confirmative of the audit results.

Evidence from Litigation

As far as I know, there is only one pending case alleging racial discrimination in new car sales, but it apparently reveals both deliberate discriminatory policies by a dealership and substantial discriminatory premiums in actual transactions.¹⁵ According to testimony by former salespersons at an Atlanta dealership, management put pressure on them to charge blacks higher prices and to offer less-favorable terms on trade-ins and financing. Evidence from sales commission sheets apparently confirms that blacks did indeed pay higher prices than whites for comparable cars. Although evidence from a single dealership is subject to the obvious question of generalizability, it does strengthen the case that discrimination in the market for new cars is an ongoing phenomenon in actual transactions, not just in unconsummated audits.

What Next? The Implications of Car Audits

Discrimination in the Car Market

Additional work on discrimination in the market for new cars could usefully take several forms. It would be interesting to replicate the audit results in another large city. Even better would be to test for discrimination using more reliable measures of transaction prices than the data Goldberg used. One strategy would be to sample dealership records, preferably from a number of dealerships. The buyer's race could be inferred from his or her residential zip code or by direct interviews. It seems unlikely that dealerships would willingly turn such information over to researchers, however. An alternative would be to start from tax records, which are publicly available in certain states, and which describe the purchase price and the make and model of car in some detail. Again, the buyer's race could be inferred from his or her zip code, or preferably through a supplementary interview with the purchaser in which additional information about bargaining and search could also be obtained. The major obstacle to this kind of research is the problem of how to oversample black new car buyers. In Goldberg's data, only 67 of 1,300 respondents (5 percent) were "minority," with blacks making up presumably only 60 percent of that group, or less.

Discrimination in Other Markets

There are at least three important differences between automobile dealerships and most of the other public accommodations contexts where audits might be deployed. First, negotiated car prices are both flexible and hidden. Discrimination in this context takes the form not of denial or degradation of service to black customers, but of charging them more than otherwise identical whites would be charged for the same product. Discrimination based on idiosyncratic bargaining presumably increases seller profits and is much easier to conceal



than outright denial of service. Other things equal, this leads me to expect more discrimination in car sales than, say, in restaurant meals.

A second important difference between the car market and other kinds of public accommodations is that the problem of incomplete transactions is not really applicable when discrimination takes the form of denials of service; in a sense, it is precisely the incomplete transaction that audits are trying to capture.

Third, issues of tester matching, which are crucial in the context of employment and housing audits, and significant in auditing car dealerships, are likely to be substantially less important in auditing restaurants or shops. Restaurant testers presumably do not need simulated biographies or extensive coaching on presentation of self; although the testers need to resemble each other in terms of dress and general demeanor, more than that does not seem necessary.¹⁶

Other Markets in Which Racial Price Discrimination May Be Prevalent

As Graddy points out, “. . . price discrimination based on race in retail markets has been recently almost ignored” (Graddy 1997, 391).¹⁷ Despite the lack of empirical evidence, table 1 does suggest that there are other markets with the same structural characteristics as the market for new cars. Consider, for example, the markets for home repairs, television repairs, used car purchases, or new appliances. In each of these markets, products or services have an important idiosyncratic dimension. For example, although a repair shop may post its hourly rate for work on a television set, most consumers are in no position to know what is actually wrong with their television, or how long it should take to fix it. Under such circumstances, sellers could easily charge different prices based on the race of the customer. Whether this actually occurs is an open question, which could be addressed relatively easily using paired audit techniques.¹⁸

Taxicabs

Prompted by repeated complaints, the Washington Lawyers’ Committee sponsored an audit study of racial discrimination in taxi service in the District of Columbia. The study design involved pairs of testers, one white and one black, who were positioned “about three car lengths apart” at selected Washington locations.¹⁹ The testers were randomly assigned to be “first” in the pair. Each then attempted to hail a taxi. If a cab stopped, testers were instructed to request rides to selected locations in the city; even when the testers were successful in hailing a cab, they were sometimes refused service to their chosen destination.

What Is the Appropriate Measure of Racial Discrimination?

The results of the study are presented in table 3, and summarized in table 4 in a way that makes them more consistent with other studies.²⁰ I think the most logical measure of the discrimination black patrons face purely on the basis of their race is simply to compare the probability of successfully hailing a cab for black and white testers, as shown in table 4.²¹ The table demonstrates that whites are about 8 percentage points (11.2 percent) more likely to be able to get a taxi to stop for them than are blacks.²² Put another way, it took blacks an



Tester 1, Tester 2	Accept, Pass (a)	Accept, Attempt (b)	Pass, Pass (c)	Pass, Accept (d)	Pass, Refuse (e)	Refuse, Pass (f)	Refuse, Accept (g)	Refuse, Refuse (h)	Total
T1 Black, T2 White	45	44	3	14	12	5	12	17	151
T1 White, T2 Black	34	54	1	1	0	16	13	23	141

Source: Calculated from Ridley, et al. (1989), table 1. Rows may not sum to totals because of rounding from percentages.

average of 5.7 minutes to get a taxi to stop for them, while it took whites 4.5 minutes; blacks had to wait, on average, 27 percent longer for a cab to stop.

What Are the Costs of Discrimination?

One reading of these data is that discrimination is not such a big deal: Black patrons simply had to wait an extra 72 seconds to hail a taxi. Valuing this time at \$12 per hour, roughly the average wage, the “cost” is a modest \$0.25. However, this calculation is inappropriate because discrimination almost surely has an important psychological dimension that is not adequately captured by a simple opportunity cost valuation such as this one.

There are alternative methods for assessing the true costs of discrimination, although they are difficult to implement and theoretically controversial.²³ But the taxicab study does raise an issue that will be important in other kinds of public accommodations audits. Simply monetizing the time or inconvenience that results from discrimination runs the risk of incorrectly making it seem like no big deal. Most people would instinctively recognize that discrimination in jobs or housing carries significant costs, but an additional 1.2 minutes spent hailing a taxi may not seem that important to some observers. Audits can tell us a great deal about the frequency of discrimination in public accommodations, but we need to be prepared with surveys or other approaches to measure its costs appropriately.

1. PR(Stop Black) ^b	0.728
2. PR(Stop White) ^c	0.809
3. Difference (2-1)	0.081
4. Percent Gain from Being White (3/1)	0.112

Source: Calculated from table 2, above.

Notes: ^aTreats refusals to transport passenger to his or her chosen destination and attempts to pick up both testers as “acceptances.”

^bPR(Stop|Black) is the probability that a cab will stop for a black tester, regardless of position and regardless of whether the cab subsequently refuses to take the fare. It is calculated from table 2 as [(Row 1: cols a, b, f, g, and h) + (Row 2: cols b, d, e, g, and h)]/(151+141).

^cPR(Stop|White) is the probability that a cab will stop for a white tester, regardless of position and regardless of whether the cab subsequently refuses to take the fare.



Are Simultaneous Paired Audits Appropriate in This Context?

Start with the following story about what motivates cab drivers. Drivers believe that blacks are marginally less profitable passengers than whites,²⁴ and other things equal, will therefore prefer to pick up a white customer. If the difference in profitability between black and white passengers is relatively small, however, drivers will usually be willing to pick up black passengers; bypassing a potential paying customer requires continued search for the next customer, and if customers are scarce, the cab remains empty while the driver continues to look for the next fare. At times when customers are plentiful (e.g., at rush hour), drivers may still find it in their interests to bypass blacks in search of more profitable whites, because they expect to find their next (white) customer relatively quickly. But—again assuming a small profitability difference—drivers will usually prefer an available black passenger to the uncertain prospect of trying to find a white alternative.

If this theory accurately describes drivers' behavior, then the paired audit with simultaneous testers seems to me to be an inappropriate way to measure discrimination, for two reasons. First, the situation being audited is not typical of the setting that black customers typically encounter. When attempting to hail a cab, the testers positioned themselves 30 to 40 feet apart; both were almost certainly visible to a cab driver who was considering which, if either, of them to stop for. In this setting, drivers might always prefer the white customer, even if, in a more realistic context—one in which the alternative to passing a black patron is an indefinite search for the next customer—the black customer might have substantially less trouble hailing a cab.²⁵

A second problem with the experimental design is more technical: The outcomes for the two auditors are not statistically independent, which means that the estimated race effect may be biased.²⁶ Consider trying to investigate the effect of fertilizer on plant growth. A control group of seeds is planted on plot A and receives no fertilizer. A treatment group is planted on identical plot B, but it does get fertilized. The difference between the average heights of the plants in the two groups is then taken to be a measure of the effect of the fertilizer.

But this is a valid procedure only if the treatment of the plants in plot B has no effect on the height of the plants in plot A. Suppose that adding fertilizer to the treated plants caused them to grow taller and cast shade on the plants in the neighboring control plot, retarding their growth. The observed difference in the average heights of the two groups of plants would then overstate the true effect of the fertilizer, because it would reflect not only the effect of the fertilizer on the treated plants, but also the (negative) effects of shade on the control group. Much the same problem could occur in the taxicab tests if, as seems likely, the presence of the white tester reduces the probability that the black tester will be able to hail a cab.²⁷

Given the possible interaction effects caused by the proximity of the two testers, it might make sense to position the testers on opposite sides of the street (instead of using two testers who stand within a few yards of each other), so that they are not in direct competition for the same cabs. Randomizing which tester is assigned to which side of the street would then allow for comparison of each tester's success rate.



In sum, the taxicab study suggests three lessons for future audits. First, it is useful to measure race discrimination as the difference (or ratio) in probabilities of service, by race. By this measure, white testers were about 11 percent (8 percentage points) more likely to hail a cab than their black counterparts. Second, audit studies by themselves cannot adequately measure the cost of this discrimination. Simply valuing the additional time that blacks spend waiting for a cab at its appropriate opportunity cost is virtually certain to understate the true cost of discrimination. Public accommodations audits will have to face this problem, because unlike housing or employment, it may not be intuitively obvious to some people that such discrimination is more than an inconvenience in this context. Third, audits should be designed to ensure that the treatment of one tester will not influence the outcome experienced by the other.

Other Kinds of Public Accommodations

I propose a broad definition of “Public Accommodations” that encompasses most of the commercial transactions of everyday life, including eating in a restaurant, renting or buying a car, hailing a taxi, or going shopping. It doesn’t matter how one defines the phrase, however,²⁸ because regardless of the definition, there is virtually no systematic evidence about the extent of discrimination in public accommodations outside of the few areas discussed earlier.²⁹ One could almost stop here: The strongest case for conducting audits in this area is simply that we are almost completely ignorant of the very thing audits are designed to measure—how pervasive is discrimination in restaurants, shopping, hotels, car rentals, and so on?

We can gain some useful—albeit tentative—insights about public accommodations discrimination from two flawed sources: (1) survey data measuring self-reports of experiences with discrimination, and (2) individual reports (including journalistic accounts and formal legal opinions) of discriminatory behavior in various kinds of public accommodations. I consider these in turn.

Survey Data

The Limitations of Survey Data

Before I discuss the survey results themselves, it is important to recognize their limitations, even though these problems do not ultimately bear on the case for audits.

Surveys of discrimination are plagued by two obvious problems. First, respondents may not be aware of some instances of discrimination. This is particularly likely when it takes the form of higher prices—as in the case of the instances of discrimination uncovered in the audits of auto dealerships, virtually none of which were recognized by the victims—as opposed to outright refusals of service. In a world where race discrimination is illegal in most contexts and is widely considered to be immoral, discriminators have both a legal and a social incentive to practice deceptive “Have a Nice Day Racism” rather than overt discrimination. When discrimination is difficult to detect, we cannot



count on victims to give an accurate estimate of its extent, and surveys are thus potentially flawed as measures of the frequency of discrimination.

But the reverse problem also exists. Respondents may incorrectly classify some instances of simple bad service (a long wait for a table at a restaurant, for example) as race-based discrimination when in fact the bad service is not racially motivated.

It is impossible to know a priori how important either of these effects actually are. But whether victims overstate or understate the amount of discrimination they face, there is a clear social benefit to a better, more objective understanding of the prevalence and nature of discrimination.

The Implications of Some Recent Survey Evidence

A 1997 Gallup study provides some of the most recent and authoritative survey evidence of blacks' experiences with discrimination (Gallup 1997, 30–31).³⁰ Asked whether they had encountered discrimination (unfair treatment because of their race) in the last 30 days, 45 percent of blacks said that they had had at least one discriminatory experience. Thirty percent said they had experienced discrimination while “Shopping;” the figure for “Dining Out” (including bars, theaters and other entertainment) was 21 percent (Gallup 1997, 30).

The Rate of Discrimination in Public Accommodations. Even assuming that these results are perfectly accurate, it is not obvious how to turn them into the appropriate rates of discriminatory behavior. But a crude calculation is outlined in table 5. Suppose 50 billion meals are served in restaurants and school and work cafeterias each year, of which roughly half are in commercial establishments, and roughly 10 percent of these are served to blacks. If the black dining public consists of 25 million persons, then the average black customer eats roughly 8 restaurant meals a month. A group of 100 black respondents should thus eat roughly 800 restaurant meals per month. Twenty percent of respondents claim they experienced discrimination at least once during the past month. Conservatively supposing that each victim experienced only a

Table 5 *A Rough Estimate of the Prevalence of Race Discrimination in Restaurant Dining*

1. Total Number of Meals Served in Restaurants and School and Work Cafeterias, <i>per year</i> ^a	50 Billion
2. Of which, Commercial Meals ^b	25 Billion
3. Of which, Served to Blacks ^b	2.5 Billion
4. Number of Black Restaurant Customers ^b	25 Million
5. Restaurant Meals per Black Customer, <i>per Month</i>	8.3
6. Percent of Blacks Who Say They Have Experienced Discrimination While Dining Out During the Past Month ^c	20 percent
7. Percent of All Meals Served to Blacks That Result in Perceived Discrimination ^d	2.5 percent

Sources and Notes:

^aSource: National Restaurant Association website (1998), <<http://www.restaurant.org/research/pocket/index.htm>>.

^bRough Estimate.

^cSource: Gallup (1997, 30).

^dAssuming that the 20 percent who reported discrimination had one discriminatory and 7.3 nondiscriminatory meals, while the remaining 80 percent of respondents had 8.3 nondiscriminatory meals, the share of nondiscriminatory meals among all meals is $(0.8 \times 8.3 + 0.2 \times 7.3)/8.3 \approx 0.975$. See text for caveats.



single incident of discrimination, we have 20 acts of discrimination out of 800 total meals, for a discrimination rate per meal served of about one in forty, or 2.5 percent. This is substantially lower than (net) rates of discrimination found in the Urban Institute's employment audits, which ranged from 5 to 15 percent.³¹

These calculations have two important implications for auditing discrimination in public accommodations. First, even if the rate of discrimination in some activity is relatively low, people who perform that activity frequently will nevertheless have a high probability of experiencing discrimination at some time.³² Consider the rate of discrimination in "Shopping." It is hard to know what the interviewers or respondents meant by the word "Shopping," but broadly defined, the average respondent probably went "Shopping" dozens of times in the month before he or she was surveyed. If so, even though 30 percent of respondents experienced discrimination on a shopping trip, the rate of discriminatory incidents *per trip* would be extremely low—on the order of 1 percent—and hence very difficult to detect via random audits.³³

Public accommodations audits thus pose a different set of challenges to investigators than housing or employment audits. One problem concerns questions of sample size and statistical power. For example, suppose that, using audits of a random sample of stores, we measure the rate of discrimination in shopping at 1 percent per trip in 1998, but that in 2003 we observe a rate of 0.05 percent, a 50 percent drop. If it were "real," such a large drop in the rate of discrimination would obviously be extremely important. But it is of course possible that the decrease could be caused simply by sampling variation: It could be that the 2003 sample just happened by chance to include a group of firms that were less likely to practice discrimination, even though the overall (population) rate of discrimination remained unchanged.

Our ability to distinguish between a real (population-level) change in the rate of discrimination and an artifact of sampling variation depends on the precision of the two estimates. This in turn depends on the size of the two samples. It is useful to ask: How big a sample size would we need to be able to reject the null hypothesis of no real change at the 5 percent significance level, given that the rate was 1 percent in 1998 and appeared to fall by 50 percent in 2003? The answer is about 4600 observations (2300 in each year).³⁴ This is dramatically larger than the biggest testing study ever conducted, and probably not feasible. If we observed a decline of less than 50 percent (which seems more plausible), we would need an even larger sample size to distinguish sampling error from a real decrease. Statistical power considerations thus make it extremely difficult to assess changes in discrimination rates over time when discrimination rates are already relatively low, as they appear to be in some public accommodations activities such as shopping.

Heterogeneity. The other important fact that emerges from the Gallup survey is that the incidence of perceived discrimination apparently varies a great deal *within* the black population. As table 6 indicates, there appear to be significant effects of age and gender, with young black males reporting substantially more discrimination than any of the other age/gender groups. It is impossible to know whether these results reflect differences in perceptive acuity, different definitions of what constitutes discrimination, or actual differ-



Table 6 *Percent of Black Respondents Who Report Having Experienced Discrimination Within the Last 30 Days, by Activity, Age, and Gender*

Activity	Men Ages 18–24	Women Ages 18–24	Men Over 35	Women Over 35
Shopping	45	28	25	26
Dining Out	32	24	19	15

Source: Gallup (1997, 31), based on a survey conducted in Jan./Feb. 1997, with a total of 1269 black respondents. (N for this table not available.)

ences in treatment; but it is interesting that at least the gender pattern observed here is consistent with the results of the Ayres/Siegelman car audits, in which black male testers were quoted dramatically higher prices than any of the other three groups. In any case, the data suggest that future audit studies need to pay attention to age and gender as well as race. This further complicates the audit process—requiring larger sample sizes, for example—but the heterogeneity evidenced in table 6 suggests that “one size fits all” audits may not give a true picture of the extent of discrimination.

Evidence from Litigation and Journalistic Accounts

A second source of information on discrimination in public accommodations is individual narratives, ranging from newspaper accounts to judicial opinions. But generalizing from such accounts—and especially from judicial opinions—to the social world from which they originate is extremely hazardous. Incidents that lead to litigation or generate substantial publicity are a tiny and nonrandom fraction of what actually goes on; tried cases are a small and unrepresentative proportion of filed cases; and tried cases are not randomly selected for opinion-writing or publication. Thus, although judicial opinions in public accommodations cases (and journalistic accounts of discriminatory incidents) are an important source of what we *think* we know about public accommodations discrimination, their message is not always what it seems.

Table 7 provides what I believe to be a reasonably complete listing of all the judicial opinions in (federal and state) public accommodations cases since 1990, plus some of the additional incidents that have received substantial press coverage. I want to consider three aspects of this table: the small number of cases, the possible importance of “Race-Plus” discrimination, and the identity of defendants.

Why So Few Cases?

Perhaps the most striking fact about table 7 is its length. Although there have been tens of thousands of federal employment discrimination cases filed since 1990, and several thousand opinions written, my search turned up a mere 23 opinions in public accommodations cases in both state and federal courts.³⁵ The contrast is striking. It is even more striking if we combine it with the estimates in table 5, which suggest that blacks experience something on the order of



Table 7 *State and Federal Court Decisions and Other Incidents Involving Race Discrimination in Public Accommodations Between January 1990 and January 1998^a*

Name	Description	Outcome
<i>Morris v Office Max</i> , 89 F.3d 411 (7 th Cir. 1996)	Black male customers incorrectly suspected of shoplifting and questioned by security guards	Plaintiffs lost on summary judgment & on appeal
<i>Alexis v McDonald's Restaurants of Mass., Inc.</i> , 67 F.3d 341 (1 st Cir. 1995)	Dispute between black customer and Hispanic clerk leading to ejection of patrons	Plaintiffs lost on summary judgment & on appeal
<i>Evans v Holiday Inns, Inc.</i> , 951 F.Supp. 85 (D.Md. 1997)	Ejection of allegedly rowdy black patrons by motel management	Plaintiffs lost on summary judgment
<i>Efstathiou v Romeo Carryouts & Liquors, Inc.</i> , 1997 U.S. Dist Lexis 14810 (N.D.Ill.)	White males accompanied by black females denied service at diner	Unclear, but plaintiffs survived motion for summary judgment
<i>Haywood v Sears, Roebuck & Co.</i> , 1996 U.S. Dist. Lexis 11954 (E.D.N.C.)	Black customers incorrectly suspected of shoplifting, interrogated/harassed by security guards	Unclear, but plaintiffs survived motion for summary judgment on some claims, though <i>not</i> on those relating to discrimination in public accommodations
<i>Perkins v Marriott</i> , 945 F. Supp. 282 (D.D.C. 1996)	Dispute over whether room rate included breakfast led to confrontation between black couple and hotel staff	Plaintiffs lost on summary judgment
<i>Lewis v J.C. Penney Co.</i> , 948 F. Supp 367 (D.Del. 1996)	Black customer accused of shoplifting, allegedly treated differently than white friend shopping with her	Plaintiff lost on summary judgment
<i>Perry v Burger King Corp.</i> , 924 F. Supp. 548 (S.D.N.Y. 1996)	Black customer allegedly denied use of bathroom because of his race	Unclear
<i>Jackson v Motel 6, Inc.</i> , 931 F. Supp. 825 (M.D.FI. 1996)	Black police officers told that motel was full; white officers subsequently obtained a room	Unclear, but plaintiffs survived motion for summary judgment
<i>White v Denny's Inc.</i> , 918 F. Supp. 1418 (D.Colo. 1996)	Restaurant allegedly seated white customers before black plaintiffs, then sided against plaintiffs in dispute with white fellow patrons	Plaintiff lost on summary judgment on all federal claims, unclear on state law claims
<i>Jackson v Tyler's Dad's Place, Inc.</i> , 850 F. Supp. 53 (D.D.C. 1994)	Black customers allegedly denied seating in restaurant because of their race	Plaintiffs lost on summary judgment
<i>Robertson v Burger King</i> , 848 F. Supp. 78 (E.D. La. 1994)	Black patron alleges whites who arrived after him were served first	Plaintiff lost for failure to state a claim
<i>Harvey v NYRAC, Inc.</i> , 813 F. Supp. 206 (E.D.N.Y. 1993)	Black plaintiff alleges she was denied a car rental because of her race	Unclear, but plaintiff survived motion for summary judgment
<i>Estiverne v Saks Fifth Ave.</i> , 1992 U.S. Dist. Lexis 18089 (E.D. La.)	Black customer whose check was not approved alleges race discrimination	Plaintiff lost on summary judgment and was subject to Rule 11 sanctions
<i>Bermudez Zenon v Restaurant Compostela, Inc.</i> , 790 F. Supp. 41 (D. Puerto Rico 1992)	Group allegedly denied seating at restaurant because some were black	Unclear, but plaintiffs survived motion for summary judgment
<i>Stearnes v Baur's Opera House, Inc.</i> , 788 F. Supp. 375 (C.D. Ill. 1992)	Black patron alleges club deliberately selected music blacks wouldn't enjoy in order to keep them out	Plaintiff lost on summary judgment

(Continued on page 84)



Table 7 *State and Federal Court Decisions and Other Incidents Involving Race Discrimination in Public Accommodations Between January 1990 and January 1998^a (continued)*

Name	Description	Outcome
<i>Franceschi v Hyatt Corp.</i> , 782 F. Supp. 712 (D. Puerto Rico 1992)	Hotel refused to allow son of black patrons to visit them on the premises	Plaintiff survived summary judgment
<i>Bray v RHT, Inc.</i> , 748 F. Supp. 3 (D.D.C. 1990)	Black patron alleges he was asked to leave restaurant because of his race	Plaintiff lost on summary judgment
<i>Roberts v Walmart Stores, Inc.</i> , 736 F. Supp. 1527 (E.D.Mo. 1990)	Black patrons object to their race being recorded on check they used to purchase items from store	Unclear, but plaintiff survived motion for summary judgment on some claims
<i>Jones v City of Boston</i> , 738 F. Supp. 604 (D. Mass. 1990)	Black patron alleges he was subject of abusive remarks by bartender at hotel	Unclear, but plaintiff lost on summary judgment on most claims
State Cases		
<i>Crook v American Inn., Inc.</i> , 680 So.2d 361 (Ala. Civ. App. 1996)	Black couple and six children denied room at hotel, despite reservation. Hotel claims "2-to-a-room" policy	Plaintiff lost at trial and on appeal
<i>Jackson v Superior Court</i> , 30 Cal.App.4th 936 (Cal. App. 1 Dist. 1994)	Black investment advisor accused by bank employee of attempting to defraud his clients	Unclear, but plaintiff's loss on summary judgment overturned on appeal
<i>Clarke v KMart Corp.</i> , 495 N.W. 2d 820 (Mich. App. 1992)	Black shopper detained/harassed by store in dispute over behavior of clerk	Unclear, but plaintiff's loss on summary judgment overturned, in part, on appeal
Other Incidents of Public Accommodations Discrimination		
Denny's Restaurants (1993–1998)	Several incidents at restaurants across the country involving disparate treatment of black customers (refusals of service, prepayment of bill, etc.)	Cases settled after suits filed. Settlement involved structural changes to corporation, expanded minority recruiting, monitoring, and \$46 million in payments to injured claimants
Shoney's Restaurants (?–1992)	Widespread corporate policy of discrimination against customers and employees	Shoney's agreed to \$100 million settlement of some 10,000 employment discrimination claims
Dillard's Department Stores (1997)	Black shopper wrongly accused of shoplifting, allegedly because of her race. Security guards apparently had explicitly race-conscious policy	Jury awarded plaintiff more than \$1 million
Avis Car Rental (?–1997)	Franchisee explicitly trained staff to avoid renting to black customers; no evidence of pattern and practice of discrimination elsewhere, although corporate HQ may have ignored complaints about this franchise	Avis and franchisee agree to \$3.3 million settlement of class action lawsuit; individual claims continue
Eddie Bauer Clothing Store (1997)	Store security personnel wrongly detained black youth suspected of shoplifting; there was apparently no race-conscious policy at issue	Jury awarded plaintiff more than \$1 million

^aThe federal cases were found in Westlaw's ALLFEDS database using the key numbers covering violations of civil rights relating to public accommodations in general; in inns, restaurants, bars and taverns; in theaters; in public conveyances; and in places of business or public resorts. I excluded discrimination in private clubs. The exact search was: (78K119 78K120 78K121 78K122 78K123) and DA(AFT 1/1/1990) and RACE. This was supplemented with a Lexis search in the Courts library using 42 USC 2000a and date aft(1/1/90). The state cases were located using the identical searches in the ALLSTATES database. The searches produced 59 federal and 37 state cases, of which only those included here were relevant.



(0.025×2.5 billion =) 60 million discriminatory incidents in restaurants alone each year.³⁶

What is going on? I think the most likely possibility is that potential plaintiffs realize that a large fraction of the perceived incidents of discrimination they experience are in some sense not worth the costs of taking to court.³⁷ This is not to suggest that such incidents are psychologically unimportant, but only that plaintiffs (or their lawyers) who perform the simplest cost/benefit calculation probably conclude that the expected monetary gains from litigation are unlikely to be greater than the costs.³⁸

If true, this scenario makes it pretty clear that we cannot rely on private citizens to enforce the civil rights laws prohibiting discrimination in public accommodations.³⁹ The case for audits as an enforcement tool requires more than this, however. We still need to know whether market forces can effectively provide a check on discriminatory behavior, and whether audits can be designed so as to detect the discrimination that does occur.⁴⁰ Sorting all this out would be a major contribution that audits could make to our understanding of discrimination and to the design of an effective enforcement effort.

“Race-Plus” and the Problem of Auditing Special Circumstances

While table 7 contains many instances in which plaintiffs allege that they were simply refused service or mistreated because of their race,⁴¹ many of the incidents being complained about involved a normal transaction that somehow went awry. In one case, black customers got into a dispute with a cashier at McDonald’s about what they had ordered; the patrons were then forcibly ejected from the restaurant, even though they were apparently eating peacefully at the time they were thrown out.⁴² In another example, a dispute over whether the room rate included breakfast led to a confrontation between plaintiffs and the hotel staff.⁴³ It is impossible to be precise, but these kinds of “race-plus” incidents, in which race combines with some other factor to generate disparate treatment, seem to account for one-fourth to one-third of the cases in table 7.⁴⁴

Let me be clear that these “race-plus” cases do not imply the absence of discrimination. Rather, they suggest that discriminatory treatment is often conditional on something else *in addition* to race. For example, suppose that black hotel guests receive the same treatment as white guests as long as there are no complaints about the service (which is equally bad for both races). If they do complain, however, black customers are then subject to hostile or disrespectful treatment that complaining whites do not receive. This clearly constitutes discrimination, but it is in some crucial respects different from the stark denial of service to black customers.

The existence of “race-plus” discrimination poses a problem for the design of audits because tests that look only at the treatment of “exemplary” customers of both races will understate the true level of discrimination when it takes the form of “nonexemplary” blacks being treated worse than nonexemplary whites. Of course, it is virtually impossible to know what proportion of all perceived instances of discrimination is accounted for by this kind of “race-plus” discrimination; in the end, it may not prove to be a major part of the problem.⁴⁵ A well-designed survey could illuminate the importance of this kind of “race-



plus” effect and would be an extremely useful precursor to designing public accommodations audits.⁴⁶

Who Is Being Sued? Another surprising fact about table 7 is the identity of defendants listed there. Sixteen of the twenty-three defendants are large, national, publicly traded corporations such as Sears, Holiday Inn, and Burger King. Moreover, all of the recent, well-publicized incidents of public accommodations discrimination seem to have involved large, national firms: Shoney’s, Denny’s, Avis Rent-a-Car, Dillard’s, and Eddie Bauer.

In spite of the pattern that seems to emerge from table 7, I would expect that discrimination is more prevalent at smaller, single-location shops and restaurants. To be sure, large national chains are capable of discriminatory conduct, either as a matter of top-down policy⁴⁷ or as an unauthorized exercise of low-level managerial discretion.⁴⁸ But economic theory strongly suggests that national chains are substantially less likely to discriminate overtly than single-outlet shops or restaurants. The reason is simply that McDonald’s, Sears, and Holiday Inns risk losing black customers at *all* their outlets, *nationwide*, if they are perceived to be discriminating against blacks at *any* individual outlet.⁴⁹ A locally owned diner that caters to highway travelers can afford to serve bad food (or to discriminate against blacks), knowing that most of its customers are “one-shotters” who will never patronize the restaurant again regardless of the quality of food or service it provides. But a bad meal or a discriminatory experience at a restaurant that relies on repeat customers for a substantial share of its business is much more costly to the restaurant. Even though any individual customer may not patronize the same McDonald’s more than once, a bad experience at a McDonald’s in Dubuque could lead customers to shun their local McDonald’s in D.C., and this possibility provides McDonald’s with a strong incentive not to provide a substandard dining experience for any of its customers.⁵⁰

Neither the cases in table 7 nor the widely covered incidents at Denny’s, Shoney’s, etc., constitute random samples of real-world behavior. Discrimination at large, nationwide entities is newsworthy in a way that discrimination at the corner gas station is not, so press reports will therefore be much more likely to ignore the latter and concentrate on the former. Larger defendants are more attractive to potential plaintiffs for a variety of reasons.⁵¹ Even though the evidence seems to suggest the contrary, I believe that public accommodations audits are much more likely to uncover discrimination at single-outlet entities than at national chains, which typically have too much to lose by encouraging or permitting discriminatory practices.

Depending on one’s goals, this analysis suggests either a testable hypothesis or an enforcement strategy. A sensible first step in either case would be to audit both large, nationwide chains and small, single-outlet entities. Enforcement efforts could then be concentrated on the latter if, as I expect, the probability of encountering discrimination is found to be higher there.⁵²

Conclusion

Discrimination has often been found in places where one might think a priori that it was impossible or unlikely. The important question about discrimination



in public accommodations is not whether it occurs at all, but how often, in what circumstances, and what can be done about it.

Audits are a necessary, but not sufficient, technique for answering these questions. They are necessary because they provide virtually the only objective measure of discriminatory treatment in many contexts, especially where discrimination takes the form of refusal or degradation of service rather than higher prices. There are simply no alternative transaction-based measures of how often race is a factor in getting service at a restaurant or while shopping.

In one sense, public accommodations audits will be significantly easier to perform than housing or employment tests because the problems of matching testers (including the creation of false biographies) are much less challenging in this context than in earlier studies. On the other hand, public accommodations audits pose some technical challenges that have either not been considered before or have received insufficient attention.

First, if my analysis of the survey data is correct, the “low-incidence/high-frequency problem” could make discrimination in shopping or restaurants difficult to detect without either a priori targeting of suspected discriminators or substantially larger sample sizes than have been used in previous testing.

Second, if “race-plus” discrimination is significant, conventional audits based on “exemplary behavior” could miss an important aspect of discriminatory behavior in public accommodations. Designing audits that can capture the effects of nonexemplary behavior may prove to be impossible for reasons discussed earlier. In that case, it is important to stress that the results of testing should be interpreted as *conditional* on exemplary behavior, and as understatements of the true amount of discrimination that minorities actually face.

A third, technical issue in interpreting the results of public accommodations audits might be termed the General Equilibrium (or “Other Margins, Other Rooms”) Problem. As I noted in discussing the car studies, if people who might experience discrimination take avoidance measures (for example, by refusing to patronize suburban malls where they might be accused of shoplifting), then audits that randomly sample all stores could well produce higher estimates of the extent of discriminatory behavior than survey data based on actual experience would. This may not be evidence against the reliability of audits, but could instead simply indicate that some of the true costs of discrimination are experienced along some margin other than the one on which it nominally occurs—as higher search costs or diminished shopping opportunities, for example.⁵³

Finally, tests for racial discrimination in public accommodations also need to be sensitive to differences of gender and age, since the survey data show such dramatic differences in (perceived) discriminatory treatment by age and gender. Social class and region could be significant as well, although the survey data do not allow for breakdowns on these dimensions. The survey results could reflect differences in treatment, or in perceptions. It is precisely because we need to sort out which explanation is correct that audits should be designed to shed light on these issues.

Audit studies, especially if they are used for enforcement, should therefore be complemented by well-designed surveys that can help reveal those areas where discrimination is most likely and help uncover the true costs of



discrimination, which may involve avoidance, higher search costs, and other alternatives that are not revealed by audits.⁵⁴

Surveys should be designed to measure rates of discrimination accurately. Rather than asking respondents if they have experienced discrimination during the last month, they should ask in detail about how many times the respondent could have been exposed to discrimination (e.g., how many restaurant visits) and on how many of those times, if any, the respondent actually experienced discrimination. Surveys should try to measure the nature of discrimination: Was it outright refusal of service? Was it delay in getting a table? They should also ask in detail about the respondent's subsequent behavior: Did he or she leave, complain, file suit? Why? Finally, it is important to know about avoidance measures that respondents may have taken: Are there stores, restaurants, or malls that respondents will not patronize in an effort to avoid discriminatory treatment?

Survey research techniques might usefully be complemented by experimental evidence. Social psychologists have developed many clever experimental techniques for uncovering the importance of race in explaining whites' "helping behavior," aggression, and nonverbal communications.⁵⁵ Many of these studies could usefully be replicated over time, supplementing survey research questions as an indicator of the evolution of white attitudes toward blacks. Although behavior in experimental settings is not direct evidence of discrimination in the real world, carefully designed experimental studies can play an important role in assessing the background level of prejudice that motivates certain kinds of discrimination.

A final word about enforcement. In an era when some elected officials have suggested abolishing the Internal Revenue Service because its audits are too intrusive, the idea of covert discrimination audits at local restaurants, movie theaters, hotels, and department stores is unlikely to be greeted with much enthusiasm.⁵⁶ I think the political realities thus argue strongly for a two-stage process. Stage one would involve purely descriptive/analytical social science research; enforcement audits would be used only if/when this research uncovers a serious problem. There is a practical reason for a two-stage approach, as well: if I am correct that shopping and restaurant discrimination is relatively uncommon, enforcement audits might need to be targeted toward those firms, industries, or regions where discrimination is most likely.

Endnotes

1. The audit studies in employment and housing are summarized and critiqued in Fix and Struyk (1993). For a summary of research in employment discrimination using conventional regression methods, see Cain (1986).
2. John Yinger (1998) surveys some of the same landscape and reaches roughly similar overall conclusions.
3. Kathryn Graddy (1997) concludes that posted prices at fast food chains are higher in minority neighborhoods, after controlling for a large number of supply-side variables such as crime rates and labor costs. But Graddy's results do not contradict the emptiness of cell 2: Fast food chains appear to be price discriminating on the basis of neighborhood characteristics, but do not charge different prices because of the race of an individual customer.



4. It will often be obvious to the potential patron that he or she has been bypassed by a vacant taxi, and in this sense, such discrimination is overt. On the other hand, it is not simply the ability to *detect* discrimination, but the ability to punish it, that discourages discriminatory behavior. And even though taxicab discrimination is observable, it is virtually impossible for patrons to take their business elsewhere or invoke legal sanctions against an individual cab driver. Hence, taxicabs could be classified in cell 3 instead of cell 4.
5. The original idea was Ayres's, and his analysis of a smaller-scale study is contained in Ayres (1991). The results of the larger study are analyzed in Ayres and Siegelman (1995).
6. See Goldberg (1996). Note that Goldberg's own interpretation of her results is that they are at odds with ours; I argue below that this is incorrect.

The issue is important for audits in other areas as well: Our confidence in audits is strengthened if the results are confirmed using other techniques. As discussed below, however, even if audits seem to reveal more discrimination than occurs in actual transactions, this does not imply that the audits are inaccurate. It could simply mean that the discrimination uncovered in an audit study operates on some margin other than the price paid by black consumers—for instance, blacks could end up paying the same price as whites, but only by being forced to search more diligently or bargain harder.

7. Selling a car is a discrete transaction that requires less knowledge about the purchaser than hiring an employee does about the applicant. Hence, issues of tester matching are less significant in this context than in tests for discrimination in hiring. We were therefore able to “mix” testers, so that A would sometimes be matched with B, sometimes with C, and so on. The importance of matching is discussed further below.
8. See, for example, Epstein (1994, 34) (“A technique of testing that leaves so many incomplete transactions cannot be an accurate replica of a functioning market.”) and Goldberg (1996, 623) (“The reported markups [in the Ayres and Siegelman study] may . . . be different from the ones realized in actual purchases of new cars.”).
9. One possible criticism of the audit findings is that we might expect completed bargains to equalize prices paid: The higher the offer at the time negotiations ended, the larger should be the subsequent concession that the dealer would be willing to make. While seemingly plausible, there is no evidence of such behavior in our data. To the contrary, black male testers started the bargaining process by receiving the highest initial offers, and dealers conceded *less* to them than to any other tester group.
10. Column 3 of the table is the standardized difference between Goldberg's estimates and those of Ayres/Siegelman. The numbers in column 3 are test statistics for the null hypothesis of no difference between the two estimates of the minority premiums: Under the null, the difference between the two estimates has a standard normal distribution. For white females and “minority” females, one cannot reject the null hypothesis that the two estimates are identical at any conventional significance level. The black male estimates are significantly different, however.
11. How should one interpret the fact that Goldberg's estimates of the discriminatory premiums were not statistically significant while ours were? Goldberg suggests that these differences in statistical significance make her results different from ours. She is unable to reject the hypothesis that her estimates were generated purely by sampling variation from a population where the true average discriminatory premium was zero. She seems to suggest, therefore, that the true discrimination premium actually *is* zero, not the much larger parameter values we both estimated.

A more convincing interpretation is simply that the differences in the statistical significance of our results arise because we are both measuring the same underlying parameter(s), but that she is doing so with data that are substantially noisier than ours. Among the factors that make it harder for Goldberg to precisely measure the discriminatory premiums are (1) she did not distinguish between various minority groups, while we employed only black testers; (2) her data are for the race and gender of the household head, not the actual purchaser of the car; (3) her data on the options purchased with each car, and even the model purchased, are less detailed than ours; (4) roughly half of her transactions involved trade-ins, whose value can be assessed only on average (using the wholesale blue-book value for a given make, model, and year); (5) Goldberg's data do not include the household's state of residence and do include sales taxes, which could only be factored out probabilistically; and (6) the survey data she used



were collected as long as three months after the actual transaction took place and are subject to errors of recall or memory that add noise to the price and other variables of interest.

Further evidence of the noisiness of Goldberg's data is that her R2s were about 0.18, while ours ranged from 0.22 to 0.44.

12. One possible explanation for the discrepancy is that the race variable in Goldberg's study is measured with error, which would impart a downward bias to her estimated race effects. More plausibly, if there are variables omitted from Goldberg's survey data that are positively correlated with race and negatively correlated with price, their omission would result in biased (understated) estimates of the race premium.
13. The importance of these kinds of general equilibrium concerns has long been recognized in the labor market context and is discussed at greater length below. See, for example, Flinn and Heckman (1983) or Duleep and Zolotar (1991). Yinger (1997) formalizes the intuition that discrimination by sellers reduces the benefits of additional search by buyers, causing them to accept higher prices or lower quality than they otherwise would. Yinger applies this methodology to the housing market and finds that the costs of discrimination are roughly \$4,000 per minority household per search.
14. Goldberg (1996, 643–44) also tests for whether minorities respond to discrimination by switching to the second-hand market or deciding not to buy a car at all; she concludes that there is no evidence for either of these effects in her data. However, there is independent evidence that, controlling for income, education, region, and age, blacks are less likely to own a car, and more likely to own an older or used car, than are whites. See Mannering and Winston (1991, table A-1).
15. See *Williams v Sutherland*, No. 1-96-CV-1215 (N.D.Ga. 1996). For a discussion of the law governing race-based price discrimination, see Ayres (1991, 857–63).
16. On the importance of matching in the context of employment audits, as well as other methodological issues, see Heckman and Siegelman (1993).
17. In other work, Graddy has found race to have a significant effect on prices at the wholesale level in the highly competitive Fulton Fish Market in New York, where buyers of Asian ethnicity paid roughly 5 percent *less* for identical fish. See Graddy (1995). Wholesale fish are relatively standardized and homogeneous in quality. But apparently, buyers are nevertheless unable to compare prices charged by the same seller to different customers. If race-based price discrimination is possible in a highly centralized market with sophisticated repeat buyers, it should be all the more likely in markets that lack such characteristics (such as television repair).
18. Many years ago, the Federal Trade Commission conducted a study in two states that found a high level of consumer fraud in the television repair industry. See Phelan (1974). Although not the same as race-based price discrimination, this fraud does suggest that such discrimination is possible in settings where consumers are unable to compare the price they pay with that paid by others.
19. More detailed results and descriptions of the methods, as well as raw materials such as testing procedures and other protocols, can be found in Ridley, Bayton, and Outtz (1989).
20. In analyzing race discrimination, the taxicab study decomposed the possible audit outcomes into several categories. A “passby” occurred when the taxi refused to stop for a tester; an “acceptance” meant that the taxi stopped *and* agreed to take the tester to his or her chosen destination; a “refusal” occurred when the taxi stopped for the tester but would not agree to go to his or her destination; and finally, an “attempt” occurred when a taxi picked up the first tester and then stopped for the second tester as well. (D.C. cabs are allowed to pick up multiple passengers, but testers were told not to accept rides from a cab that had already stopped for their partner.)

The authors chose to compare the rate of “passbys” among “those passby outcomes where race might be a factor” with the rate of “acceptances.” But a case can be made for including both “refusals” and “attempts” along with “acceptances” for these purposes. If the taxi was willing to stop for the tester, but chose to refuse him or her only when told of the tester's destination, this might better be thought of as discrimination on the basis of destination rather than race.



Given patterns of residential segregation, discrimination on the basis of the customer's race and on the basis of his or her destination are likely to be highly correlated. But there is a difference: One can think of a taxi that refuses to stop for a black tester as committing a kind of disparate treatment discrimination; a taxi that refuses to go to a (mostly black) neighborhood is committing a kind of disparate impact discrimination. Although this distinction is in one sense academic, I think it is one that is worth preserving because it could have important policy implications.

21. The probability of hailing a cab includes all of the tests in which a black tester was able to hail the cab, including those in which his or her white partner had already hailed it and those in which the cab refused to take the tester where he or she wanted to go. By focusing only on columns d, e, and f, the taxicab study authors ignore the fact that nearly 40 percent of the time, a taxi that picked up the white tester first attempted to pick up the black tester as well.
22. The difference is statistically significant at the 0.02 level using a χ^2 test with 1 d.f.
23. Note that there are two kinds of costs of discrimination that need to be measured. The first are the opportunity costs imposed on victims who have to bear increased search costs or who substitute away from one activity to what would otherwise be a less-preferred alternative in order to avoid discrimination—for example, taking a bus instead of a taxicab. These kinds of costs are discussed by the authors cited in note 13.

But there are almost certainly additional costs that are purely psychological and therefore more difficult to measure. One possible measurement strategy would be to use a “willingness to accept” approach. Suppose everyone were guaranteed nondiscriminatory taxi service. Now imagine (as only an economist could) asking black customers what's the least amount of money it would take to get them to give up their right to nondiscriminatory service, if this meant they would have to wait an additional 72 seconds, on average, to hail a cab? The answer to this question is one measure of the true cost of discrimination. My intuition is that it would be dramatically larger than the 25-cent figure derived above.

So-called “contingent valuation” methods, based on surveys of this sort, have been widely used in attempting to value nonmarket goods such as environmental quality. They are controversial within the economics profession, and I don't want to suggest that they are the definitive answer to the problems of assigning a cost to discrimination. For a positive assessment of the methodology, see Hanemann (1994); a critical view is offered by Diamond and Hausman (1994).

24. There could be many reasons for such a belief. Drivers might think that they will have a harder time finding a return fare from predominantly black neighborhoods that are the likely destinations of black passengers; or that black passengers are more likely to rob them; or that blacks leave lower tips than whites. Some evidence for this last possibility comes from an unpublished study by Ian Ayres and Suzanne Perry of Yale Law School, which finds that black taxi patrons left significantly lower tips than whites, controlling for destination and length of trip.
25. Suppose that drivers' motivations were based purely on animus toward blacks, rather than on profitability concerns. The results should be no different. Following Becker (1971), drivers' animus can be modeled as an implicit psychological tax that prejudiced drivers incur when they pick up black passengers. If the tax is low relative to the costs of searching for another customer, drivers will prefer to pick up the black patron and pay the “tax,” rather than continuing to search for a preferred white passenger. If an alternative white customer is immediately available, however, the driver will have no reason to incur the psychological costs that result from serving a black customer and will choose to pick up the white patron instead.
26. In this context, statistical independence means that $PR(\text{Taxi Stops} \mid \text{Customer Is Black})$ is the same as $PR(\text{Taxi Stops} \mid \text{Black Customer and Alternative White Customer 40 Feet Away})$. If the two conditional probabilities are not the same, the presence of the second tester influences the probability of the first tester's successfully hailing a cab. Contrast this with testing in the context of housing, employment, or car purchases. The large numbers of applicants or buyers should mean that the presence of one tester has no effect on the outcomes of the others.

Note that nonindependence also implies that the standard errors used to construct test statistics are also incorrect.



27. Note that nonindependence is much less likely to be a problem with the housing or employment audits discussed in Fix and Struyk (1993) or new car audits conducted by Ayres and Siegelman (1995). In the housing and employment context, testers were instructed to turn down job offers or apartment rentals in order to prevent a successful outcome from influencing the success or failure of their audit-mate. In the new car study, one tester was unlikely to influence the price quoted to his or her partner because the number of prospective buyers who visit a showroom without buying is so large.
28. Of course, enforcement of laws prohibiting discrimination *does* require a precise definition, and courts in the heyday of the Civil Rights era did have to struggle with these kinds of definitional issues, especially in interpreting Title II of the 1964 Civil Rights Act, 42 U.S.C. § 2000(a)-(a)6, which prohibits discrimination in various public accommodations. But most of these issues have been settled for decades; the last public accommodations case heard by the Supreme Court was apparently in 1973 (*Tillman v Wheaton-Haven Recreation Assn., Inc.*, 410 U.S. 431).
- Essentially, Title II forbids racial discrimination in hotels, motels, inns, restaurants, gas stations, movie theaters, and other places of entertainment. Section 1981 of the 1866 Civil Rights Act, 42 U.S.C. § 1981, prohibits (intentional) race discrimination in the making and enforcement of contracts; it presumably applies to most instances of discrimination in shopping.
29. This is not to deny the existence of discriminatory conduct, of which the record contains numerous examples—Shoney’s, Denny’s, and so on. But the plural of anecdote is not data.
30. The poll was based on telephone interviews with a representative sample of 1,269 blacks and 1,680 whites who were interviewed in early 1997. The margin of error for a percentage estimate for blacks is approximately ± 5 percentage points. Gallup (1997, 5–6).
- There appears to be very little social science literature that speaks directly to the question of how frequently blacks encounter discrimination in public accommodations. Wilson (1980) asserted that poverty, rather than discrimination, is the real obstacle blocking the advancement of the black urban underclass. Wilson’s book does not squarely address the basic question of how much discrimination is actually out there, however. Feagin (1991) asserts that discrimination is still important, but his data are useless for estimating how frequently it occurs. More recently, Thernstrom and Thernstrom (1997) skirt this question by focusing not on the current prevalence of discrimination but on its rate of change over time—allegedly negative.
31. This calculation is not sensitive to the assumption about how restaurant meals are distributed among the population: It wouldn’t matter if 20 percent of the respondents did all of the restaurant dining and the remaining 80 percent did none. But the calculation *is* sensitive to the total number of meals consumed and to the assumption that those reporting discrimination experienced only a single incident. Fewer total visits or more incidents per respondent would obviously raise the rate of discrimination per meal; the relationship is linear, so doubling the number of meals or discriminatory incidents per person doubles the estimated rate of discrimination per meal.
32. In fact, under the above assumptions, the probability that a black person will experience discrimination in a restaurant at least once during the course of a year is $(1 - (1-.025)^{100})$, or 92 percent.
33. If we assume one shopping trip per respondent per day, then by the same logic as in table 5, the probability of discrimination per shopping trip is roughly 1 percent.
34. The test statistic is $(O-E)/SE$, where O is the observed difference in discrimination between the two years, E is the *expected* difference if the null hypothesis (no change in the rate of discrimination) is true, and SE is the standard error of this difference. The test statistic has a standard normal distribution. Assume that the observed rate of discrimination in year 1 is 1 percent and that it is half as big (0.5 percent) in year 2. Hence, we have $O = 0.01 - 0.005 = 0.005$. By hypothesis, $E = 0$.

The standard error of the difference is $SE = \sqrt{\sigma_1^2 + \sigma_2^2}$, where the subscripts denote the first and second years being compared. To calculate σ_i , take $\sqrt{p_i(1-p_i)/N_i}$, where p is the probability of encountering discrimination. Thus, $\sigma_1 = \sqrt{.01 * .99/N_1}$; $\sigma_2 = \sqrt{.005 * .995/N_2}$. To simplify, assume $N_1 = N_2$.

The critical value of the test statistic at the 5 percent significance level is 1.96. Hence, we are looking for values of N such that $(O-E)/SE \geq 1.96$. Given that $O-E = 0.005$, we need SE



less than 0.00254, which in turn requires $N \geq 2285$. Notice that the identical problem occurs if we try to compare discrimination rates across regions of the country, instead of across points in time.

35. Another important question that might arise from table 7 is why plaintiffs seem to be winning so small a fraction of the cases. It is hard to be precise about what constitutes a win, and much relevant information is missing, but plaintiffs clearly failed to prevail at a very early stage of adjudication in more than half (13 of 23) of the cases listed. Does this imply that meritorious claims of public accommodations discrimination are rare, and hence that discrimination is not really a serious problem in this context? The answer is clearly “No.” The reason is that the cases that are adjudicated to the point at which an opinion is written are not a random sample of all filed cases, let alone all potential suit-generating incidents. Without knowing how many cases settled, and on what terms, we should not infer anything from the results of adjudicated cases.
36. If public accommodations cases generate opinions at the same rate as employment discrimination cases, then these 23 opinions should represent something like 160 filed cases. But unlike Title VII (before the 1991 Civil Rights Act), some of the public accommodations statutes allow jury trials, which rarely generate published opinions. Settlement rates may also be higher in public accommodations than in employment discrimination cases. Both factors mean that there were probably more than 160 cases filed, but it is hard to say more than this.
37. Incidents of discrimination generally have low rates of legal claiming compared with other kinds of grievances, in large part because discriminatory intent is often difficult to prove. See Miller and Sarat (1980); Curran (1973). But see Kritzer et al. (1991) (discrimination grievances have a higher rate of claiming than traditionally thought). Moreover, litigation is a costly process, and people do not usually sue over a single rude remark or a few minutes’ delay in being seated at a restaurant. Even though such incidents can be psychologically very wounding, the costs of litigation make it an impractical response to most perceived instances of discrimination.

Another important consideration is the damages a potential plaintiff could expect to receive if he or she prevailed. Although “courts are in general agreement that punitive damages may be awarded in appropriate circumstances in actions to recover for violations of civil rights statutes” (Annotation, *Punitive Damages in Actions for Violations of Federal Civil Rights Acts*, 14 A.L.R. FED. 608, § 2a (1998)), it seems unlikely that they would routinely be granted for an isolated incident. Actual damages might include some payment for emotional distress or other psychological harm, but it would presumably also be small in most cases.

In sum, potential plaintiffs simply may not have an incentive to enforce their legally protected rights, given the high costs of doing so and the meager returns they could expect to receive. Class action litigation offers a partial solution to some of the structural problems posed by large numbers of small monetary injuries, but class action public accommodations suits apparently are relatively rare.

38. For compelling accounts of the psychological impact of discrimination while shopping, see Buchanan (1997). The existence of the phrase “shopping while black”—with its allusions to “driving while intoxicated” or other criminal offenses—suggests that the problem of discrimination while shopping is frequent enough to merit its own shorthand description in the black community.
39. This is not to suggest that the laws are unimportant. It is obvious that legal penalties against discrimination played a major role in abolishing segregated eating and recreational facilities in the decade after 1964. See, for example, *U.S. v Boyd*, 327 F.Supp. 998 (S.D.Ga. 1971) (supplemental decree) (requiring detailed changes in operation of Vandy’s Bar-B-Q in Statesboro, Ga., in order to eliminate segregation); *Katzenbach v McClung*, 379 U.S. 294, 296 (1964) (applying Title II of the 1964 Civil Rights Act to Ollie’s Barbecue, “a family-owned restaurant in Birmingham, Alabama, specializing in barbecued meats and homemade pies. . . [that] has refused to serve Negroes in its dining accommodations since its original opening in 1927”).
40. It is easy to overlook the role of market forces in recent celebrated cases of public accommodations discrimination such as those involving Denny’s or Shoney’s. Litigation or the threat of litigation undoubtedly played a major role in the extensive reform efforts that both these corporations undertook; but adverse publicity, the potential loss of business, and even pres-



sure from the capital markets also apparently played an important role, albeit not until the discriminatory practices had been exposed. On Shoney's, for example, see Kerr (1993) and *Chicago Tribune* (1992) (replacement of CEO and possible return to discriminatory practices caused stock price to fall).

41. See, for example, *Jackson v Motel 6, Inc.*, 931 F.Supp. 825 (M.D.Fl. 1996) (black police officers told that motel was full; white officers subsequently obtained a room); *Harvey v NYRAC, Inc.*, 813 F.Supp. 206 (E.D.N.Y. 1993) (black plaintiff allegedly denied a car rental because of her race).
42. *Alexis v McDonald's Restaurants of Mass., Inc.*, 67 F.3d 341 (1st Cir. 1995).
43. *Perkins v Marriott*, 945 F.Supp. 282 (D.D.C. 1996).

44. The analogy is to so-called "Sex-Plus" discrimination in employment discrimination law, under which an employment policy that does not discriminate solely on the basis of sex, but on sex *and* some other neutral classification, may nevertheless be held to violate Title VII. See, for example, *Phillips v Martin Marietta Corp.*, 400 U.S. 542 (1971) (employer who refused to hire women, but not men, with pre-school-age children is potentially liable under Title VII); *Willingham v Macon Telegraph Publ. Co.*, 507 F.2d. 1084 (5th Cir. 1975) (employer who refused to hire men, but not women, with long hair not liable under Title VII because the "plus" characteristic did not involve a fundamental right).

The difference between "race-plus" and "sex-plus" discrimination is that the latter involves an official policy that openly treats men and women differently (women, but not men, can have long hair). In "race-plus" discrimination, there can be no overt policy to treat the two groups differently, since a policy that whites, but not blacks, can protest about how long they've had to wait for a table would clearly be illegal. "Race-plus" discrimination could result from covert policies or could emerge out of some combination of fear, prejudice, or stereotypes.

45. It may be relevant that John Donohue and I noticed a similar pattern in our study of employment discrimination cases, in which frequently "a worker is fired . . . because of some alleged individual misconduct such as tardiness. The worker then alleges that those of the opposite race or gender were either less productive or even more guilty of the alleged offense but were not fired." Donohue and Siegelman (1991, 1012) (extensive citations omitted).

Zwerling and Silver (1992) reached a similar conclusion after an extensive review of the careers of 2,100 newly hired workers in one post office district. Even though they concluded that most of the fired black workers in their study probably deserved to be fired, they found a significant disparity in black/white firing rates, which they attributed to the fact that many whites who deserved to be fired were not. Thus, the discrimination they observed was based not on race alone, but on race *plus* some other factor (e.g., unexcused absence from work).

46. "Race-plus" audits pose some significant challenges. Ordinary audits are taxing enough, but asking the auditors deliberately to start a confrontation with shopkeepers or hotel clerks in order to engineer a "Race-plus" situation seems both difficult and dangerous. Experiments involving confrontation are possible in very controlled circumstances, as evidenced by the work of Nisbett and Cohen (1996), which compared reactions of northern and southern students when a confederate of the experimenter bumped into the subject in a school hallway and called the unsuspecting subject an "asshole." But they seem too dangerous to attempt in a field setting. (On the other hand, there may be special circumstances that could generate "race-plus" discrimination without confrontation: for example, testers could attempt to return goods without a store receipt or could show up one minute after closing time and ask to be admitted to a store.)

Moreover, even a survey could run into the general equilibrium problem discussed earlier: Respondents may not report experiencing many instances of "race-plus" discrimination precisely because they believe (whether correctly or not) that it is a serious problem and therefore always feel constrained to act on their best behavior in any public accommodations setting. Note that the reverse could also be true: suppose black customers draw the reasonable, but sometimes incorrect, inference that when they receive bad service, it is because of their race. This reaction, especially when it is incorrect, could provoke precisely the hostile or surly counterreaction that constitutes "race-plus" discrimination.



47. The Shoney's case appears to be a textbook example of racist management that discriminated against blacks, both as employees and as customers, as a matter of corporate policy.
48. A North Carolina Avis franchise strongly discouraged its personnel from renting cars to blacks; but as far as I know there is no evidence to suggest that this was corporate policy or that other Avis franchises also practiced this kind of discrimination. In fact, subsequent testing at other Avis outlets apparently revealed no irregularities.
49. Moreover, an explicit national policy of discrimination is more difficult to conceal because it requires more participants, at least one of whom is likely to find it objectionable.
50. In fact, McDonald's (and presumably many other large corporations) already has its own internal auditors who pose as customers, order meals, and note any shortfalls from company standards, including incorrect orders, slow service, and so on. Hostile treatment of black customers would presumably be detected by such internal audits and met with the appropriate sanctions. This analysis follows closely on Nelson's underappreciated article (Nelson 1976).

Of course, profit-maximizing managers might conclude that discriminating against blacks encourages more white customers than the black business it forecloses, as Shoney's apparently believed. But the careful survey evidence in Sniderman and Piazza (1993) suggests that there are unlikely to be enough white racists to make this kind of discriminatory strategy profitable. This is especially true when the possible legal penalties are factored in.

51. A short list might include deeper pockets, greater stakes in appearing not to be discriminating (higher willingness to pay), more personnel (and therefore a higher chance of finding a disgruntled insider who can testify to discriminatory practices), more customers (so that a pattern or practice of discrimination is easier to detect), and so on.
52. Of course, enforcement efforts have to take account of the benefits, as well as the costs, of enforcement. A lawsuit or settlement involving Joe's Diner might affect the treatment of 1,500 black customers each year, whereas getting a Denny's to change its practices could easily result in improved service for 100 times that number of patrons.
53. The flip side of this problem is part of what Orlando Patterson has called "The Ordeal of Integration": If minorities now feel freer to go places and do things they formerly avoided, they may be more likely to run into a racist even as the percentage of discriminators declines. In this sense, incidents of discrimination can be increasing even as the number of discriminators is decreasing. See Patterson (1997, 52–63). John Donohue and I similarly argued that increasing workforce integration makes it easier to *detect* discrimination, since it gives black workers a more accessible white co-worker against whom they can measure their treatment; hence, integration can lead to more employment discrimination suits at the same time that the amount of discrimination is falling. See Donohue and Siegelman (1991, 1011–14).
54. Surveys can also help explicate the psychological costs of discrimination, which, I suspect, do not necessarily track the dollar amounts at stake. In part because it is covert, discrimination in new car sales probably has a much smaller psychological impact on black buyers than does discrimination in restaurants or shopping, even though the dollar amounts at stake are much larger.

It is also important to acknowledge the social costs of discrimination, which, again, audits are not well suited to uncover. Perceived discrimination presumably results in alienation, loss of faith, and disaffection that undermine the social fabric in subtle but potentially important ways.

55. For an excellent review of such studies, now unfortunately rather dated, see Crosby, Bromley, and Saxe (1980).
56. In fact, generalized random auditing of public accommodations could easily give rise to charges of "Discrimination Nazis run amok" and so on. Rather than an Occupational Safety and Health Administration model of general inspection, a more sensible enforcement strategy would rely on audits only when a complaint of discrimination has already been initiated from some other source.



References

- Ayres, Ian. 1991. "Fair Driving: Gender and Race Discrimination in Retail Car Negotiations." *Harvard Law Review* 104:817–72.
- Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *American Economic Review* 85:304–21.
- Becker, Gary. 1971. 2nd ed. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Buchanan, Jean. 1997. "Black Citizens Tell Tales of Bias: In Wake of Dillard's Verdict, Other Shoppers Share Their Experiences." *Kansas City Star*, 14 December, A1.
- Cain, Glen. 1986. "The Economic Analysis of Labor Market Discrimination: A Survey." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and Richard Layard, 693–785. Amsterdam: Elsevier, Scientific Publishers.
- Chicago Tribune*. 1992. "Shoney Stock Falls on News of Shakeup." 22 December, B3.
- Cosby, Faye, Stephanie Bromley, and Leonard Saxe. 1980. "Recent Unobtrusive Studies of Black and White Discrimination and Prejudice: A Literature Review." *Psychological Bulletin* 87:546–63.
- Curran, Barbara. 1973. *The Legal Needs of the Public*. Chicago: American Bar Foundation.
- Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* 8(4) 45–64.
- Donohue, John J., III, and Peter Siegelman. 1991. "The Changing Nature of Employment Discrimination Litigation." *Stanford Law Review* 43:983–1032.
- Duleep, Harriet Orcutt, and Nadja Zalokar. 1991. "The Measurement of Labor Market Discrimination When Minorities Respond to Discrimination." In *New Approaches to Economic and Social Analyses of Discrimination*, edited by Richard Cornwall and Phanindra Wunnava. Westport, Conn.: Greenwood, Praeger.
- Epstein, Richard A. 1994. "Standing Firm, on Forbidden Grounds." *San Diego Law Review* 31:1–56.
- Feagin, Joe. 1991. "The Continuing Significance of Race: Anti-Black Discrimination in Public Places." *American Sociological Review* 56:101–16.
- Fix, Michael, and Raymond Struyk, eds. 1993. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: Urban Institute Press.
- Flinn, Christopher, and James J. Heckman. 1983. "Are Employment and Out of the Labor Force Behaviorally Distinct States?" *Journal of Labor Economics* 1:65–93.
- Gallup. 1997. The Gallup Poll Social Audit on Black/White Relations in the United States. 10 June. <<http://www.gallup.com/poll/special/race.htm>>.
- Goldberg, Pinelopi Koujianou. 1996. "Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* 104(3):622–54.
- Graddy, Kathryn. 1997. "Do Fast Food Chain Prices Discriminate on the Race and Income Characteristics of an Area?" *Journal of Business and Economic Statistics* 15:391–401.



- _____. 1995. "Testing for Imperfect Competition at the Fulton Fish Market." *Rand Journal of Economics* 26:75–92.
- Hanemann, W. Michael. 1994. "Valuing the Environment through Contingent Valuation." *Journal of Economic Perspectives* 8(4):19–43.
- Heckman, James J., and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael Fix and Raymond Struyk, 187–258. Washington, D.C.: Urban Institute Press.
- Kerr, Rhonda. 1993. "Shoney's: 'A Changed Company.'" *Nashville Business Journal*, 20 September, 1.
- Kritzer, Herbert M., et al. 1991. "To Confront or Not to Confront: Measuring Claiming Rates in Discrimination Grievances." *Law and Society Review* 25:875–87.
- Mannering, Fred, and Clifford Winston. 1991. "Brand Loyalty and the Decline of American Automobile Firms." *Brookings Papers on Economic Activity: Microeconomics*, 67–103.
- Miller, Richard, and Austin Sarat. 1980. "The Emergence and Transformation of Disputes." *Law and Society Review* 15(3–4):525–65.
- Nelson, Robert. 1976. "The Economics of Honest Trade Practices." *Journal of Industrial Economics* 24(4):281–93.
- Nisbett, Richard, and Dov Cohen. 1996. *Culture of Honor: The Psychology of Violence in the South*. Boulder, Colo.: Westview.
- Patterson, Orlando. 1997. *The Ordeal of Integration: Progress and Resentment in America's "Racial" Crisis*. Washington, D.C.: Civitas/Counterpoint.
- Phelan, J. 1974. *Economic Report: Regulation of the Television Repair Industry in Louisiana and California*. Staff report to the Federal Trade Commission. Washington, D.C.: Federal Trade Commission.
- Prime Time Live*. 1992. "True Colors," 26 November. ABC Television.
- Ridley, Stanley, James A. Bayton, and Janice Hamilton Outtz. 1989. "Taxi Service in the District of Columbia: Is It Influenced by Patrons' Race and Destination?" Washington, D.C.: The Washington Lawyers' Committee for Civil Rights under the law. Mimeographed.
- Sniderman, Paul, and Thomas Piazza. 1993. *The Scar of Race*. Cambridge, Mass.: Harvard University Press.
- Thernstrom, Stephan, and Abigail Thernstrom. 1997. *America in Black and White, One Nation, Indivisible*. New York: Simon and Schuster.
- Wilson, William J. 1980. *The Declining Significance of Race*. Chicago: University of Chicago Press.
- Yinger, John. 1998. "Evidence on Discrimination in Consumer Markets." *Journal of Economic Perspectives* 12(2):23–40.
- _____. 1997. "Cash in Your Face." *Journal of Urban Economics* 42:339–65.
- Zwerling, Craig, and Hillary Silver. 1992. "Race and Job Dismissals in a Federal Bureaucracy." *American Sociological Review* 57:651–60.



Chapter 5

Minority Business Development: Identification and Measurement of Discriminatory Barriers

TIMOTHY BATES

Introduction

Minority business enterprises are growing and expanding into new industries where minority presence has historically been minimal. Attraction of highly educated, experienced minorities into small-business ownership has facilitated this progress. More sophisticated owners, other things equal, have greater access to capital sources for financing business formation and growth. The more successful minority businesses today also typically compete in the broader economy—often selling goods and services to large corporate and government clients. The traditional niches for minority-owned businesses—such as shop-keeping concentrated in minority residential areas and serving local clienteles—have been in a continuous state of decline for more than 30 years (Bates 1993).

This progress notwithstanding, minority businesses are still underrepresented in many sectors and tend to be smaller and less viable than their majority counterparts. This chapter reviews existing evidence on discriminatory barriers to the formation and expansion of minority-owned businesses in order to discuss how best to improve our knowledge of the extent of such barriers.

Measurement Challenges in Quantifying Discriminatory Barriers

Three fundamental measurement challenges face any effort to quantify discriminatory barriers facing minority-owned businesses. The first is to focus on minority ownership subgroups that face similar types of discrimination. The second is to select the appropriate subgroup of nonminority firms for any particular comparison. The third is to choose industries, stages in a firm's life, and the types of barriers to be measured in any particular study. All these selection decisions have important bearing on the likelihood of finding discrimination, if it exists; the degree of bias in the estimates; and the degree of policy-relevant explanatory power of the findings.

I begin by reviewing briefly each of these measurement challenges to provide context for my discussion of what is known about discriminatory barriers facing minority businesses and how we can fruitfully learn more by using various methodologies, including audit studies.

Selection of Ownership Subgroups

Because the community of minority-owned businesses is not a homogeneous group in any meaningful sense, focusing on subgroups that face the same types of barriers is important to ensure that measured differences in treatment can be correctly interpreted. If one group experiences one type of discrimination and another group experiences a totally different type of discrimination, the true story for each group is likely to be obscured.

Thirty years ago, when President Nixon was promoting black capitalism, the terms “minority-owned” and “black-owned” were practically synonymous. But in the nation's largest urban areas today, most of the minority self-employed are not black Americans. Immigrant-owned small businesses now outnumber those owned by nonimmigrant minorities in urban America. In the Chicago metropolitan area, for example, 56 percent of minority businesses are immigrant-owned. Asian-owned firms are more numerous there than black businesses, and 93 percent of the Asian owners are immigrants.¹

Both African-American and immigrant small businesses face barriers to success, but their situations are very different in ways that must be kept clear when designing studies to measure discrimination. Asian-American immigrants, for example, are more likely than whites to be college graduates. They also devote, on average, greater financial resources to their small business startups than whites. Their problems stem primarily from the fact that many lack fluency in English. This language barrier, combined with employer reluctance to recognize foreign credentials, prevents many from finding suitable salaried employment and pushes them toward self-employment (Min 1993; Min 1996).² Those least fluent in English (Koreans) are precisely the ones most likely to own small businesses—leading to the crowding of highly educated Asians with substantial capital into low-yielding lines of self-employment, such as retailing, which in turn lead to lower average self-employment earnings among Asians than among whites (Bates 1997a). The problems for African Americans provide a sharp



contrast. Many would like to pursue self-employment but lack the human and financial capital access often available to their white counterparts.

Choosing the Majority Comparison Group

The essence of measuring disparate treatment is a comparison between groups that enables enough variables to be held constant between groups to make clear statements about differences. In this respect, firm size is a particularly crucial variable to hold comparable in studies of majority and minority firms. Differential treatment has plausibly left minority firms smaller than majority firms. But small firms' problems are different from the problems facing large firms for a variety of reasons having nothing to do with race or ethnic group. To abstract the effect of race from the effects of these other factors, the most appropriate comparison is between minority and majority small businesses. The importance of size comparability is discussed further later in this chapter. In addition to being composed of comparably sized firms, the majority comparison group obviously should match the minority business group in the industry and stage of business development relevant to a particular line of inquiry.

Choosing Industry, Stage of Business Development, and Type of Discrimination

Choosing industry is important because discriminatory barriers operate differently in different areas. Table 1 summarizes industry concentrations among minority-owned and white-owned small businesses that were active in 1992.³ Overall, 11.5 percent of the nation's small businesses are minority owned (panel A). Minorities are overrepresented in nonskilled services (particularly personal services; see panel B). These businesses areas have low capital requirements and typically serve a neighborhood clientele. Minority-owned businesses can, therefore, form and grow to a point without having to penetrate access barriers either to financial capital or to nonminority markets.⁴

In skilled services, minority-owned businesses are slightly underrepresented, at 10.5 percent. These include professional services; business services; and finance, insurance, and real estate. Minority gains in these fields have been widespread since the 1960s, arguably in good part because affirmative action has reduced discriminatory barriers in higher education, with concomitant rapid growth in minority representation in the skilled services industry (Bates 1997a).

In the manufacturing and wholesale goods industry, minority businesses are considerably more underrepresented, accounting for 8.8 percent of small firms. This is an area that requires substantial capitalization—a fruitful area, therefore, for investigating discriminatory barriers in accessing capital.

In the construction industry, minority-owned firms represent an even smaller percentage of small firms—8.3 percent. This is an area where insider networks significantly shape access to work. “Beneath the complicated regulations and proliferation of collective bargaining contracts lies a different reality, one dominated mainly by personal contacts and informal networks” (Waldinger



Table 1 Major Industry Groups in Which Small Businesses Nationwide Were Operating, 1992

	Majority-Owned Firms	Minority-Owned Firms	Total
A. Industry Concentration across Broad Racial Groups			
Construction	91.7%	8.3%	100%
Goods: Manufacturing and Wholesaling	91.2	8.8	100
Skilled Services*	89.5	10.5	100
All Industries	88.5	11.5	100
B. Industry Concentration within Broad Racial Groups			
Construction	12.4	8.7	
Goods: Manufacturing and Wholesaling	7.3	5.5	
Skilled Services*	38.5	34.6	
(subtotal)	58.2	48.8	
All Other Service Industries	16.1	18.2	
Retail	15.2	18.3	
Other Industries	6.9	9.9	
Industry Unknown	3.6	4.8	
Total	100	100	

*Includes business and professional services, finance, insurance, and real estate.
 Source: U.S. Bureau of the Census Characteristics of Business Owners database.

and Bailey 1991). Nonminorities entering construction self-employment often have immediate access to networks that provide them with work, whereas minority firms must build up the necessary contacts gradually during their early years of operation. Many fail in this effort. In the New York–area construction special trades, for example, 79.7 percent of majority-owned small businesses operating in 1987 were still in business at the end of 1991, compared with 58.4 percent of their minority-owned counterparts.

Choosing life-cycle stage for a particular investigation is almost as important as choosing industry, because barriers operate differently in the three major stages of business operation: formation, growth, and maturity. At the points of formation and early rapid growth, for example, businesses in capital-intensive fields need access to financial capital. As they mature, they need to expand their client base into increasingly nonminority markets. This suggests that studies measuring discriminatory barriers in financing might most effectively concentrate on studying young firms (in operation for five years or less, for example), whereas studies focusing on access to large markets might do better to concentrate on firms that have passed their initial formation and growth spurt (in operation for more than five years).

Finally, identifying specific types of barriers for the research focus is important to ensure that the observed differences between minority- and majority-



owned small business experiences can be fully explained. As noted earlier, African-American small businesses have historically encountered differential treatment in three major areas: access to human capital, access to financial capital, and access to majority markets. Perhaps the most important of these is human capital acquisition. The quality work experience that was not available, the skills acquisition process that was biased against minority applicants—these have been major shapers of America’s black-owned business community. They translate into the firm that was never formed, as well as the stunted firm that is handicapped by the owner’s lack of relevant skills and work experience. Evidence strongly suggests, notwithstanding the advances alluded to above, that there are still major discriminatory barriers to the acquisition of the human capital needed to succeed in many business areas. This type of discriminatory barrier is not well measured at the level of the firm, however, and is therefore beyond the scope of this chapter.

Access to financial capital and to major markets can be effectively addressed at the firm level, as I discuss below.

Financing Firm Creation and Operation

I restrict the discussion in this section to black-owned businesses because, as already noted, there are substantial differences between black- and immigrant-owned businesses with respect to capital available for business formation. These differences make it inappropriate to group them together when investigating discriminatory access to financing.

In 1992, the Roper Organization polled 472 black business owners across the country to gauge how they view their own firms, as well as black business generally. Asked why there were so few black-owned firms in the nation, 84 percent responded that “Black-owned businesses are impeded by access to financing.” Asked to identify major problems constraining operations of their own firms, owners most commonly replied “obtaining sources of capital” and “access to credit” (Carlson 1992).

Table 2 shows actual startup financing and sales figures for representative recently formed black-owned and white-owned firms operating nationwide in the two most capital-intensive lines of small business—manufacturing and wholesaling. All of the firms included were formed between 1979 and 1987 (i.e., young firms) and were using paid employees in 1987 (Bates 1997a). In 1987, sales revenue for the black-owned group averaged just under \$400,000

Table 2 <i>Manufacturing and Wholesaling: Young Firms Active in 1987</i>		
	African American	Nonminority
1987 Sales Revenues, Mean	\$394,208	\$1,005,884
Total Startup Capital, Mean	\$37,571	\$92,935
Leverage (Debt Divided by Equity), Mean	0.96	1.41

Source: U.S. Bureau of the Census Characteristics of Business Owners database.



THE URBAN
INSTITUTE

compared with \$1,000,000 for their white-owned counterparts. The corresponding difference in startup capital was \$37,500 versus \$93,000.

With respect to the startup phase, bank financing is by far the most common borrowing source for small businesses. In my study of small business formation nationwide, which compared white- and black-owned firms formed over the 1979–87 period, I found that 34.4 percent of the white owners and 28.8 percent of the black owners used borrowed capital to launch their small-business ventures. Among those borrowing firms, average capitalization was \$74,237 for the white group and \$35,842 for the black group. The majority of these borrowing firms—66 percent of the white firms and 59 percent of the black businesses—used debt capital borrowed from financial institutions. These figures imply that the behavior of financial institutions is likely to be a large part of the story. (Loans from family were a distant second as a borrowing source.)

Other research confirms that accessing bank financing appears to be easier for white- than for black-owned firms. With respect to loan approval, Ando (1988) found that black would-be borrowers were less likely than whites to have their business loan applications approved, and that these approval rate differences persisted even when various measures of business (and owner) credit risk were added as controls to the econometric analysis. But business loans are just one form of bank credit. My study of business startup financing (Bates 1997b) found that blacks are more likely than whites to finance business formation with consumer credit—in the form of home equity loans, credit cards, and the like. Blacks are less successful than whites in these borrowing areas also. According to Getter (1998), blacks and Hispanics are much more likely to have their consumer credit applications turned down than whites, even when applicant income, net worth, credit history, age, current monthly debt payment obligations, and self-employment status are controlled for. He also found that self-employed black and Hispanic applicants are more likely to be turned down than their non-self-employed counterparts—a difference that he did not find for white consumer credit applicants.

Once approved, the next step in the process is loan amount determination. The evidence strongly suggests differential treatment at this stage also. When I used Characteristics of Business Owners (CBO) data from the Census Bureau to investigate financial institution loan amounts received by small business startups, looking first at firms active in 1982 and then at firms active in 1987, I found two major recurring patterns. First, the average white loan recipient borrows more than twice as much as the average black loan recipient. Second, the average white loan recipient more effectively leverages his or her equity. Nationwide, among firms active in 1987, for example, the mean debt/equity ratios for white and black business startups tapping financial institution credit were 2.99 and 2.38, respectively. Controlling for borrower demographic traits, borrower human capital, firm traits, and borrower equity investment in the firm, among other factors, did not change the fundamental picture: Blacks received smaller loan amounts than whites with identical measured traits (Bates 1993; Bates 1997b).

Beyond startup financing, very different borrowing patterns may emerge. The importance of supplier credit, for example, undoubtedly increases greatly as firms beyond the startup phase seek to finance their steady-state operations.



Construction firms often start out without using borrowed capital. Yet the ability of small construction firms to take on large jobs hinges critically on the willingness of their suppliers to advance them goods on credit. There are as yet no systematic quantitative studies of supplier treatment of black- as opposed to white-owned construction firms because of lack of data. Established retail operations are also likely to rely on supplier credit for financing their normal inventory operating needs. Data on supplier financing of black- and white-owned small firms in this area are also unavailable for representative firm samples.

Credit approval and credit terms must also be included in a comprehensive investigation of financing patterns. At a minimum, the hypothesis of unequal credit access should be investigated regarding loan maturity and loan borrowing costs as well as loan approval and loan amount.

Even investigating all the above issues will not exhaust the topic of discrimination in small-business finance. Equity financing is not discussed here, in part because fewer than 2 percent of small-business startups use equity sources beyond the owner's household net worth and the wealth holding of other family members, including parents. Black-owned businesses have less access to venture-capital financing than white-owned firms with identical measured characteristics, however (Bates and Bradford 1992), and black representation among large-scale small businesses is certainly shaped by issues of access to venture capital. Yet no large-scale database permits comparisons of venture-capital utilization among ongoing white- and black-owned small businesses.

Bonding is another variant of the loan approval process in the small-business world. The ability of black-owned construction firms to take on large-scale jobs, for example, is heavily influenced by their access to bonding. Examination of black-owned small business involvement in government procurement programs reveals that local governments improve the performance of such vendors when they provide bonding assistance to the firms that actively sell to governments (Bates and Williams 1995). The area of bonding assistance and its accessibility (and cost) to black- and white-owned construction firms is an important subject that has rarely been systematically studied, largely because of database constraints.⁵

Gaining Access to Large Markets

Small businesses most commonly sell their goods and services to households. Yet as businesses mature, selling to other businesses and to government purchasers becomes very important for growth and stability. The market for selling to other businesses is an enormous one, with 43.7 percent of white-owned and 30.2 percent of minority-owned small businesses nationwide selling at least some goods or services to other businesses.⁶ The small majority-owned businesses selling to other firms are at the larger end of the scale, reporting mean 1987 sales of \$320,362, nearly triple the average sales revenue of firms selling nothing to other businesses. Those selling to other businesses often sell to government as well—25.4 percent (versus 6.9 percent of the firms selling nothing to other businesses).



The market for selling goods and services to government is considerably smaller than the business buyer market. Nevertheless, it has received much greater attention in the 1990s. During the early and mid-1980s, many large cities and some states began operating substantive, large-scale preferential procurement programs to increase minority involvement in selling to government clients. Minority businesses responded enthusiastically, signing up in vast numbers on government vendor lists and substantially increasing their bidding for government contracts.

In 1989, minority businesses accounted for 44.5 percent of the firms who put themselves on the vendor list for construction business with the city of Chicago (Getzendanner, Castillo, and Davis 1990). Evidence such as this has been widely used in studies showing high minority business availability combined with low minority vendor use by government. Such studies—which seek to justify preferential procurement on the basis of the vast disparity between minority-owned businesses’ availability to sell to government and their relatively small share of the government procurement business—are commonly called disparity studies. They are so controversial that they merit some discussion here because they pinpoint an important methodological issue concerning potential discrimination against minority businesses when they move beyond their traditional household markets.

George LaNoue, in particular, has spearheaded a powerful attack on the disparity concept as a rationale for preferential procurement programs benefiting minority-owned businesses. “Ignoring company size constitutes a major methodological flaw in disparity studies,” he notes (1994). According to this argument, most minority-owned firms have zero employees, and it is not realistic to assume that such tiny firms have the ability to compete for government procurement contracts. “Most disparity studies ignore firm size and include as equally available for even the largest, most complex contracts every business in the Census or on some list, whether it is a part-time casual activity or a multinational corporation,” continues LaNoue in the same published piece. In other words, government cannot be expected to use minority businesses that are incapable of selling to public-sector clients. LaNoue further argues that the lack of capacity he says typifies minority-owned businesses makes preferential procurement programs into preferential treatment of minorities, thus violating the equal protection clause of the 14th Amendment.

Obviously, this argument is oversimplified at best, given that smaller size is one of the results of the presence of the discriminatory barriers previously discussed. Thus, old-boy networks constrain the ability of minority-owned construction firms to get work. Smaller construction firms are the result. And the government attempts to remedy this discriminatory barrier by adopting a preferential procurement program seeking to increase access to work for minority firms—which would have been unnecessary in the first place had those firms the same capacity as their white-owned competitors.

La Noue’s observations suggest the following hypothesis: that lower penetration of minority businesses reflects not discrimination but smaller average firm size, higher frequency of young firms, and a distribution across industries that emphasizes the types of businesses that government (and other businesses) are unlikely to buy from (such as beauty parlors). Is it possible that programs



sponsored by corporate America, like the National Minority Development Supplier Council, amount to special treatment of minority business suppliers in the same way that affirmative action is accused of doing in the field of minority contracting? Perhaps the fact that 30 percent of the nation's minority businesses were found to be selling to other businesses reflects a practice of giving them unfair advantages—such as sheltered markets and bid preferences—that discriminate against white-owned firms.

The key is that low minority presence does not, in and of itself, indicate discrimination. This is an important point and deserves appropriate examination. But in the recent debate over minority business capacity and involvement in preferential treatment efforts, there has been little analysis of what that capacity really is. The appropriate question, which has not been asked, is this: Among two firms that are the same age and size, and operating in the same industry, does the minority business have a smaller or greater chance of selling to other firms? Using nationwide representative samples of minority- and majority-owned firms drawn from the CBO database mentioned earlier, I address this issue econometrically in table 3, which shows the independent contribution of race, controlling for firm size, age, industry, and owner gender. In this formulation, a positive coefficient on minority business ownership

Table 3 *Delineating Firms That Sell to Other Businesses in 1987 from Firms That Do Not Sell to Other Businesses: Logistic Regression*

	Regression Coefficient (Standard Error)	Variable Mean (in tens of thousands of dollars)
Constant	.118* (.023)	—
1987 Sales	.001* (.000)	19.881
Young Firm	-.069* (.022)	.317
Minority Owner	-.610* (.020)	.087
Male Owner	.173* (.021)	.756
Construction	.073* (.032)	.140
Goods: Manufacture and Wholesale	1.253* (.029)	.072
Retail	-.608* (.027)	.170
Skilled Services	-.527* (.027)	.304
n	52,996	
-2 Log L	65,808	
Chi Square	7,640.8	

*Statistically significant at the .05 level.

Source: U.S. Bureau of the Census Characteristics of Business Owners database.



THE URBAN
INSTITUTE

would support the argument of preferential treatment.⁷ The coefficient is actually negative, however, indicating that minority-owned businesses have a much smaller chance of selling to other firms, even if size and other factors are held constant. My earlier discussion of minority firms being limited in obtaining work by entrenched networks (using black firms in construction as my example) is the crux of the matter.

The Ford Motor Company is a case in point. In 1996, Ford purchased \$2 billion of goods and services from minority suppliers. According to Renaldo Jensen, director of minority supplier development at Ford, the most difficult part of integrating minority suppliers into Ford's operations was the initial step: "getting buyers within the company to accept new suppliers" (*Minority Supplier News* 1997). In other words, status quo networks have traditionally not included minority businesses, and established networks tend to be resistant to change. This is the problem that the National Minority Development Supplier Council is designed to reduce in corporate America.

The logistic regression analysis in table 3 yields a strong finding that is inconsistent with the hypothesis that low minority-business representation in this market segment reflects lack of capacity. It supports the hypothesis that entry barriers tend to keep out minority businesses because of their minority status, other things equal. Further, the same analysis identifies lines of small business that are most and least likely to sell to other firms. Those in the manufacturing and wholesale goods industries are most likely to sell to other firms. Those in retailing are least likely to do so. Thus, an analysis of market access in selling to other firms would logically focus much more heavily on the minority-business manufacturer and wholesaler than on the minority-business retailer.

Given the size and comprehensiveness of the CBO database, it is possible to mine this data source much more deeply for information on market niches and their accessibility to various minority business subgroups. Industry-specific subgroups, for example, could explore whether construction poses particular difficulties—or, alternatively, whether the recent rapid growth in areas such as business services has made that sector more amenable to minority penetration.⁸ This potentially rich area of investigation has been little exploited, however, with use of large-scale representative minority business samples (with white firm comparison groups) particularly lacking.

The Potential for Audit Studies to Investigate Discrimination in Small-Business Finance

In judging the feasibility and cost-effectiveness of audit studies as a way to measure differential treatment, it is important to bear in mind that such studies are expensive and—unless substantial previous work has refined the analytic questions to be asked and the specific areas of activity and stages in the business process to concentrate on—are unlikely to yield findings with clear enough policy implications to be worth the cost. An area that seems particularly fruitful to me, given the work that has already been done, is small business access to finance.



Existing studies provide guidance in how to structure a pilot study of minority-business borrowing, and our knowledge to date suggests that measurement of discrimination in business finance through auditing could yield substantial advances in what we know. It would be somewhat complicated, however, because use of paired testers is unlikely to be feasible. For small businesses, creditworthiness is shaped by the traits of both the owner and the firm, yielding too many variables to permit straightforward matching of loan applicants. Applicable business traits include industry in which the firm operates, age of firm, size (annual sales volume), physical locations (more than one location is possible), firm balance sheet (particularly regarding liquidity), profitability and cash flow, collateral, and credit rating. Relevant owner traits include education, skills, work experience, gender, age, wage and salary income, self-employment income, other income, credit history, personal net worth, and collateral.

I suggest limiting initial testing to firms in operation (with the same owner) for less than five years, single-owner firms, single-location firms, and firms operating in industries where small-business borrowing is frequent. Black and white owners, all of whom possess at least five years of work experience, have attended at least one year of college, and are between the ages of 25 and 50, would be good candidates for a pilot study of discrimination in lending. This still leaves far too many explanatory variables applicable to creditworthiness to permit paired tester use, however. A solution is to draw two small-business samples that constitute a *pair of groups that are comparable on average except for race*. Econometric techniques such as logistic regression analysis can then be employed to use the testing results to develop measures of differential treatment of black and white business borrowers that control for individual differences among firms, thus simplifying both the testing and the interpretation of the findings.

Since we know that more business startup financing comes from financial institutions than from all other debt sources combined, lending by financial institutions is the primary area to test for differential treatment. The findings of pilot studies using testers to investigate discrimination in home mortgage lending suggest that it is important to test for discrimination at the preapplication stage (Galster 1993). Small-business borrowers are most commonly seeking either term loans or lines of credit when they inquire about bank loan availability. At this preapplication stage of inquiry, borrowing in any form may be discouraged, and this may be more prominent among black than white borrowers as the mortgage lending tester evidence suggests. It is also possible that a potential borrower may be steered to a government-guarantee form of loan—typically a term loan carrying a default guarantee issued by the Small Business Administration (SBA)—or a form of consumer credit. We already know that black business borrowers disproportionately use SBA loans (Bates 1984) and consumer credit (Bates 1997b) to finance small businesses.

A pilot study that looked solely at black-white potential borrower treatment at the preapplication stage of small business lending would be valuable in and of itself. It may also prove feasible to add to the value of an audit study by moving on to the actual process of filing loan applications. One complication here is that use of fabricated data to support loan applications is often regarded



as a criminal offense, thus ruling out consideration of paired testers that are made seemingly identical by means of invented characteristics. A combination of pairing representative business subgroups and econometric techniques applied to real application data can get around this problem. The ensuing analysis should then focus on differential black-white treatment in loan approval and, for approved loans, loan dollar amount, loan interest rate, loan maturity (in months), loan type, and collateral requirements.

As noted earlier, research to date indicates that black business borrowers receive smaller loans than whites possessing identical measured characteristics. Yet this finding tells us little about *how* the loan application and approval process differs for white and black business borrowers, and it is too broad to guide enforcement efforts seeking to reduce black-white differential treatment in this area. Audit studies are a promising way to fine-tune our understanding of bank small-business lending practices.

Endnotes

1. These firm counts are based upon businesses that filed a federal small-business income tax return and reported gross sales revenues of at least \$5,000 in 1992, according to unpublished tabulations by the author from the Characteristics of Business Owners data collected by the U.S. Census Bureau. Figures published by the Census Bureau on comparable firm populations include those grossing \$500 or more in sales revenues. Tabulations reported in this study were calculated on-site at the U.S. Bureau of the Census Center for Economic Studies. These tabulations do not reflect views of the Census Bureau or its Center for Economic Studies.
2. This influx of well-educated, capital-rich minority immigrants—particularly Asians—suggests the need for policymakers to rethink their minority business assistance strategies.
3. Active firms, by definition, include only those described in endnote 1. Figures published by the Census Bureau on comparable firm population include those grossing \$500 or more in sales revenues. Applying the higher \$5,000 gross revenue cutoff obviously results in a much smaller business universe than that of the Census Bureau. The reason for the higher cutoff is to distinguish small business from casual self-employment.
4. With respect to financial capita, about 30 percent of black-owned businesses (and about 24 percent of white-owned businesses) start up with zero financial capital (Bates 1997a).
5. Simple modification of the CBO questionnaire would be sufficient to generate the data necessary to permit comprehensive analysis of small-business access to bonding.
6. These proportions of firms selling to other businesses, and the small firm sales figures cited in the following sentence, were calculated by the author, using the CBO data described in notes 1 and 3 for firms active in 1987.
7. The dependent variable equals 1 if the firm sold any goods or services to other firms in 1987, zero otherwise. The explanatory variable, 1987 sales, is expressed in tens of thousands of dollars (19.881 = \$198,881). The explanatory variable, young firm, equals one for firms in operation for less than three years, zero otherwise. The other variables have self-explanatory names; for example, construction = construction companies.
8. A more detailed description of using econometric models (like that shown in table 3) to explore market access appears in Bates (forthcoming).



References

- Ando, Faith. 1988. "Capital Issues and Minority-Owned Businesses." *The Review of Black Political Economy* 16(4, Spring):77–109.
- Bates, Timothy. Forthcoming. "Viewing Minority Business Assistance as a Job-Creation Strategy." In *Job Strategies for the Black Worker in the 21st Century*, edited by Edward Montgomery. Washington, D.C.: Joint Center for Political and Economic Studies.
- _____. 1997a. *Race, Upward Mobility, and Self-Employment: An Illusive American Dream*. Baltimore: Johns Hopkins University Press.
- _____. 1997b. "Unequal Access: Financial Institution Lending to Black- and White-Owned Small Business Startups." *Journal of Urban Affairs* 19(4):487–95.
- _____. 1993. *Banking on Black Enterprise*. Washington, D.C.: Joint Center for Political and Economic Studies.
- _____. 1984. "Small Business Administration Loan Programs." In *Sources of Financing for Small Business*, edited by Paul Horwitz and R. R. Petit, 211–39. Greenwich, Conn.: JAI Press.
- Bates, Timothy, and William Bradford. 1992. "Factors Affecting New Firm Success and their Use of Venture Capital Financing." *The Journal of Small Business Finance* 2(1, Fall): 23–38.
- Bates, Timothy, and David Howell. 1998. "The Declining Status of African American Men in the New York City Construction Industry." *Economic Development Quarterly* 12(1, Feb.): 88–100.
- Bates, Timothy, and Darrell Williams. 1995. "Preferential Procurement Programs and Minority-Owned Businesses." *Journal of Urban Affairs* 17(1, Winter):1–17.
- Carlson, Eugene. 1992. "Black Entrepreneurship: Battling Bias." *The Wall Street Journal*, April 3.
- Fix, Michael, and Raymond Struyk, eds. 1993. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: Urban Institute Press.
- Galster, George. 1993. "Use of Testers in Investigating Discrimination in Mortgage Lending and Insurance." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael Fix and Raymond Struyk, pp. 287–334. Washington, D.C.: Urban Institute Press.
- Getter, Darryl. 1998. "Do Lenders Evaluate Applicants Differently?" Paper presented to the National Economic Association annual meeting, Chicago, January.
- Getzendanner, Susan, Ruben Castillo, and Milton Davis. 1990. *Report of the Blue Ribbon Panel to the Honorable Richard M. Daley*.
- LaNoue, George. 1994. "Standards for the Second Generation of Croson-Inspired Disparity Studies." *The Urban Lawyer* 26(3, Summer):521–31.
- Min, Pyong Gap. 1996. *Caught in the Middle: Korean Merchants in America's Multiethnic Cities*. Berkeley: University of California Press.



_____. 1993. "Korean Immigrants in Los Angeles." In *Immigration and Entrepreneurship: Culture, Capital, and Ethnic Networks*, edited by Ivan Light and Parminder Bhachu, 185–204. New Brunswick, N.J.: Transaction Publishers.

Minority Supplier News. 1997. "Ford Purchases from MBEs Reach a Record \$2 Billion." Winter.

Waldinger, Roger, and Thomas Bailey. 1991. "The Continuing Significance of Race: Racial Conflict and Racial Discrimination in Construction." *Politics and Society* 19 (3):291–323.



Chapter 6

The Future of Civil Rights Testing: Current Trends and New Directions

RODERIC V.O. BOGGS

Introduction

Since the Urban Institute's last testing conference in 1991, there have been significant developments in the utilization of this technique as a means for investigating discriminatory conduct and enforcing the nation's civil rights laws. During this period the use of testing has expanded dramatically in the field of housing and significant progress has been made in the refinement of testing techniques for application in the fields of employment and public accommodations. Courts and government agencies have increasingly accepted testing as a research and enforcement tool. The purpose of this section is to suggest some of the factors that should be considered in shaping a strategy for the use of civil rights testing in the future, including in particular, suggestions of new areas where testing might usefully be employed. Before these topics are addressed, it may be helpful to provide an overview of the testing activities that have been undertaken over the past few years by public and private agencies.

Fair Housing and Lending

Fair housing has been by far the area where civil rights testing has been most actively employed as a technique for research and enforcement. Although testing came into broad acceptance and expanded use following the Supreme Court's 1982 decision in the *Havens* case,¹ during the past several years there has been a dramatic increase in the overall amount of testing being conducted and the range of practices subjected to scrutiny. These expanded efforts have benefited greatly from the cooperative efforts of the private fair housing community, operating primarily through the National Fair Housing Alliance, the Office of Fair Housing and Equal Opportunity at the Department of Housing and Urban Development, and the Civil Rights Division of the Department of Justice.

Private fair housing groups have been in the forefront of national efforts to develop testing for more than 40 years. The *Havens* decision, by judicially recognizing the standing of testers and fair housing organizations to bring suit on the basis of testing evidence, opened the door to a significant expansion in the scope and effectiveness of private enforcement of the country's fair housing laws. In the years immediately following this decision, an increasing number of private housing groups began to employ testers as part of their programs.

The work of private fair housing groups gained momentum in 1988 with the establishment of the National Fair Housing Alliance (NFHA). The Alliance is a consortium of private fair housing agencies that came together to further their mutual interest in promoting equal housing. From an original membership of 30 local organizations, NFHA has grown dramatically in recent years and now serves more than 90 affiliates, operating with an annual budget of more than \$1.5 million. Through publications, conferences, and training sessions, NFHA plays a critical role in national fair housing advocacy and promotion of testing.

Among the Alliance's numerous accomplishments, perhaps none is more significant for purposes of testing than its critical role in supporting the establishment by the Department of Housing and Urban Development (HUD) of the Fair Housing Initiative Program (FHIP). Established in 1990, FHIP was designed to provide federal funding to private and public fair housing enforcement agencies. Much of this funding has been devoted to testing, both general audit testing and complaint-based testing. From an initial funding level of \$3 million in 1990, this program grew to a level of \$26 million in FY1995 and HUD has proposed a budget of \$29 million for FY1999. Private fair housing groups estimate that FHIP funding supported 7,000 housing tests nationally in 1990, a number that increased to 20,000 in 1996.

The bulk of FHIP funding has been directed to rental testing, with limited funding in the area of sales. HUD has provided major funding to NFHA on two occasions to conduct national investigations of possible discrimination in the area of homeowners' insurance and in 1993 funded a national investigation of mortgage lending practices.² FHIP funding has been enormously helpful in the creation and growth of new private fair housing groups and the expansion of established agencies. The work of the Fair Housing Council of Greater Washington over the past few years provides an excellent illustration of what FHIP funding can accomplish. The Council has used FHIP funding to test for



discrimination affecting numerous protected categories in a variety of contexts. This testing has combined elements of outreach, education, and enforcement. The results of Council tests, released periodically as part of a Fair Housing Index, have received a great deal of public attention and have been major factors in generating increased local government and community support for fair housing advocacy.

It should be noted that \$10 million in proposed FHIP funding for FY1999 is intended to support a series of audit-based enforcement tests to be conducted in 20 metropolitan and nonmetropolitan areas across the country. The results of these tests will be part of an effort to measure discrimination on an ongoing basis, with the hope that it will be possible to demonstrate over time the impact of enforcement on reducing levels of bias.

While the efforts of private fair housing groups supported by HUD have provided the preponderance of fair housing testing, the Housing Section of the Justice Department's Civil Rights Division has undertaken groundbreaking work in this area as well. The Justice Department (DOJ) began a testing program in December 1991. Over the past six years, the DOJ has used testing evidence generated by its program in nearly 50 cases. Of these cases, nearly 80 percent have been won at trial or settled on favorable terms for the government. One case was lost in court and the remainder are pending. A total of over 1,000 tests have been conducted by DOJ involving several hundred tested entities. DOJ testing has utilized several types of testers: Some tests have used non-attorney DOJ employees serving on a released time basis, while others have been done under contract with private fair housing groups and on occasion with individuals. Most of the DOJ testing has focused on issues of racial and national origin discrimination in rental practices, but testing has also been employed in connection with accessibility to new construction for people with disabilities and, in one case, the admissions practices of a nursing home. DOJ testing is distinguished from the testing done by most private groups because it is generally conducted with the aid of audio-tape recording equipment. This type of evidence has proven extremely powerful in enforcement actions.

While HUD, DOJ, and private fair housing groups have had the most extensive experience with testing in general and most lending testing has been conducted by NFHA or private agencies receiving HUD funding, both the Office of the Comptroller of the Currency (OCC) and the Federal Trade Commission (FTC) have also used testers in conjunction with their regulatory responsibilities. Both of these agencies also participate as part of the federal government's Fair Lending Task Force. An informal body established in 1994, its membership includes the various federal agencies with jurisdiction over the lending practices of banks or other financial institutions.

Over the past two years, the OCC has tested eight institutions selected primarily on the basis of Home Mortgage Disclosure Act (HMDA) data.³ The OCC tests have been conducted under contract by private fair housing groups. To date, OCC testing has not led the agency to conclude there was a basis to believe discrimination had occurred. The FTC has used testing in conjunction with its enforcement of the Equal Credit Opportunity Act.⁴ Over the past 14 years, the FTC has completed several hundred tests, the great majority of which have been conducted by telephone rather than through on-site visits by testers. The prac-



tices challenged by the FTC have included such things as the improper exclusion of child support, alimony payments, and retirement income from calculations of assets and income in determining creditworthiness, and the illegal consideration of the credit applicant's age, race, and national origin or marital status. The FTC has brought a total of 20 cases using testers, all of which have been settled.

The NFHA and the Department of Agriculture are now exploring the use of testing in connection with home mortgage lending activities of the Department of Agriculture (DOA). They hope to begin a pilot program this year to develop a methodology for testing for discrimination in rural areas affecting home mortgage loans provided by the DOA's Rural Housing Service. This effort is especially important because relatively little housing testing has been conducted in rural areas.

The overall picture presented by the experience of testing in the areas of housing, lending, and insurance over the past six years is very positive. Testing has gained broad acceptance and its use has been expanded to include new issues beyond real estate sales and rentals. Substantial progress has been made in creating a strong network of private fair housing agencies, working at the national level through the NFHA. The federal agencies with enforcement authority in the field are working well together and the primary agencies with regulatory and enforcement authority are in regular contact with one another. Although differing on some aspects of strategy and tactics, the private and public agencies working in the field share a strong commitment to promoting high technical standards in the quality of testing and expanding its use.⁵

In recent years the use of housing testing has been expanded to cover a range of new areas and protected classes. Homeowners' insurance and mortgage lending testing are primary examples of this new work, as are the efforts now under way in the field of disability access. Perhaps most noteworthy is the fact that housing testing has achieved very broad acceptance as a basic and powerful technique for documenting the existence of discriminatory conduct.

Fair Employment

Over the past nine years, there has been steady progress in the development of testing for purposes of investigation and enforcement in the field of fair employment. This effort began with the Urban Institute's pilot research testing work and the creation of the Fair Employment Council of Greater Washington in 1990. Shortly thereafter, the Legal Assistance Foundation in Chicago also began to do pioneering work in the field. Each of these organizations has contributed significantly to the development of equal employment opportunity (EEO) testing methodology and its application to civil rights enforcement. A number of notable developments have occurred since the last Urban Institute Conference on Testing in 1991.

Perhaps the most important development for future application of EEO testing is the success of the Fair Employment Council (FEC). Over the past seven years, the FEC has conducted a series of significant research studies refining the use of testing as it affects several categories of individuals protected by our civil



rights laws, and initiated two landmark cases that led to judicial recognition of testing as a means of enforcing federal and local employment discrimination laws.⁶ It has also operated as a national clearinghouse for EEO testing information. As part of this work, the FEC sponsored a national conference on enforcement testing in 1993, which introduced EEO testing to a number of federal, state, and local officials. It has continued to play a critical role in educating federal, state, and local government agencies on the potential of testing.

The litigation initiated by the FEC produced the first federal- and state-level appellate decisions on the standing of private organizations using testers to bring suit under federal and local civil rights laws.⁷ The federal case brought by the FEC against the local franchise of a major national employment referral agency upheld the standing of the private agency to sue on the basis of testing evidence and suggested that under the 1991 amendments to Title VII testers would have standing to sue on their own behalf as well. The second case involved both testers and a nontester plaintiff who alleged that she had been sexually harassed in the course of seeking job referral counseling. The appellate decision in this case affirmed the standing of the FEC and its testers to bring suit under the D.C. Human Rights Act and damages were awarded to all of the plaintiffs. The D.C. Circuit case is especially important because it was decided by a particularly conservative panel of judges. The FEC cases are also important because they were supported by amicus briefs submitted by the EEOC.⁸

The enforcement testing work of the FEC has been complemented by the efforts begun in 1991 by the Legal Assistance Foundation (LAF) in Chicago. The LAF is a not-for-profit legal service provider serving low-income clients. Its principal support comes from the federal Legal Services Corporation. Over the past seven years it has used private grant funding to develop a testing program that conducted a substantial number of tests and filed a dozen EEOC charges based on tester-generated evidence. Several cases have been filed in federal court, although none has yet produced a legal ruling on issues of standing. The LAF has developed an excellent working relationship with the Chicago Regional Office of the EEOC. As part of this relationship the EEOC regularly informs charging parties with claims of hiring discrimination of the availability of testing services from the LAF. The LAF and EEOC are now developing a pilot program that will allow the EEOC to share data on employment practices, which should lead to targeted-hiring testing of industries where minority representation is especially low.

Over the past two years, cooperative efforts involving private groups using employment testers and federal agencies have begun to expand significantly. In 1996, the Department of Labor's Office of Federal Contract Compliance Programs (OFCCP) contracted with the FEC to use testers to investigate possible employment discrimination in entry-level jobs in four industries doing business with the federal government in the Washington, D.C., metropolitan area. The results of this pilot program strongly support the adoption of testing as an enforcement tool by the Department of Labor. The OFCCP is now working with the FEC to develop a program for using testers in examining possible discrimination on the basis of citizenship and national origin in the Chicago area. This effort is being undertaken in conjunction with the Justice Department's Office of Special Counsel, which has jurisdiction over charges of discrimination based



on citizenship. At the same time, the EEOC has entered into contracts with the LAF and FEC to conduct pilot testing projects of employment discrimination in several regions of the country. The LAF recently completed work on a contract to provide training to EEOC District Office personnel on testing methodology and procedure. Although limited in scope, these federal agency initiatives are extremely important and have great potential for the future.

Recent Innovations on the Uses of Testing

Although the vast preponderance of civil rights testing in recent years has occurred in the areas of housing and employment, innovative work has been completed in several other fields as well, and ideas for testing in new areas are under consideration.

Two examples of this work are the studies of discrimination in the Washington, D.C., taxicab industry and the testing of automobile dealerships in Chicago (Ridley, Bayton, and Outtz 1989; Ayers 1991). Both of these studies revealed significant disparities in the treatment of customers based on race and in the case of the automobile sale testing additional disparities based on gender. The taxicab testing in the District of Columbia is especially noteworthy because it led to a significant decision in a case brought by testers upholding the liability of taxicab companies for the actions of their drivers.⁹ The eventual settlement of the cases provided for additional testing under the auspices of the D.C. Taxi Cab Commission.

Testing for purposes of monitoring compliance with court-approved settlements is a prominent part of the injunctive provision in the settlement of two major national cases involving Denny's Restaurants.¹⁰ These cases, one in California and the other in Maryland, involved allegations that Denny's engaged in a concerted policy of discriminatory service directed at African-American customers. Both cases were settled in 1994 for total monetary payments of more than \$45 million. Among other things, the consent decrees in these cases called for the creation of an independent civil rights monitor, who is authorized to administer hundreds of tests annually for several years to monitor Denny's adherence to nondiscriminatory customer service.

The use of testing in the Denny's settlement has a parallel in the growing practice of private fair housing groups to establish agreements with real estate companies where the private organizations provide testing services as part of voluntary compliance programs in house sales and rentals. Fair housing groups in Washington and Cincinnati have begun to operate this type of program with considerable success.

Among the most significant developments affecting testing in the past several years has been its use by the media to investigate and publicize discriminatory practices. Testing using hidden microphones and cameras was a prominent part of several network television programs in the past two years. One of these examined the treatment of African-American and white testers in seeking employment and housing and in general consumer situations. Another program examined issues affecting wheelchair users seeking housing, employment, and access to public accommodations. These programs, both of which



were produced with the assistance of private testing organizations, conveyed an enormously powerful message regarding the prevalence and emotional impact of discrimination.

Areas for Future Testing

There are many areas in which current civil rights testing for research and enforcement efforts should be expanded and a number of new areas where use of testing should be considered. Clearly, work in the area of fair housing needs to remain a central focus of testing for the foreseeable future. Although rental testing has proven highly effective in documenting discriminatory treatment and securing substantial relief for large numbers of individuals, it seems clear from audit testing that high levels of discrimination persist in housing rentals. Virtually all knowledgeable observers believe discriminatory conduct based on race and national origin has become more subtle, and therefore more difficult to detect through the simple forms of testing. There is broad agreement that a meaningful reduction in rental discrimination will require a great deal more testing and enforcement. These tests will, over time, require the use of more sophisticated testing techniques and necessitate a higher level of expenditure. More sophisticated testing techniques will also be needed in the fields of mortgage lending and homeowners insurance. Future efforts in all these areas must be expanded with the recognition that testing well beyond the pre-application or application stages will often be required. The availability of sufficient funding, especially from HUD, is a key element in ensuring that momentum is maintained.

In considering priorities for future testing, special consideration should be given to efforts to open housing opportunities for low-income families. For this reason, the testing for discrimination affecting families with Section 8 vouchers should be followed closely. It will also be appropriate to consider the development of national or regional cases, possibly involving coordinated testing efforts of multiple public and private agencies. The implications for testing related to electronic data transmission and the Internet must also be examined carefully.

Employment testing has vast potential and requires development on many levels. The recent decision of the EEOC to fund pilot studies in two regions with the FEC and Legal Assistance Foundation and the OFCCP's planned work on citizenship testing are important steps that will hopefully lead to even greater commitments of resources in the future. EEOC testing efforts to date have correctly focused to a large degree on entry-level positions in job categories employing significant numbers of new employees, especially those requiring relatively low levels of education. Coordinated efforts involving the use of labor market and census data will be extremely useful in identifying areas for future testing.

In exploring the development of employment testing, it is important to recognize that federal government support for this work is likely to develop slowly, with initial support directed primarily to audit and research testing. In light of this reality, foundation support is vital to the success of efforts to build on the judicial precedents established through the pioneering work of the FEC.



Looking beyond housing and employment there are many other areas where testing shows great promise, but to date has not been employed to any significant degree. The experience with taxicab and restaurant testing illustrates the potential for the technique for investigating complaints and monitoring settlements. As further demonstrated by the FTC's experience, testing can be productively applied in regard to many economic transactions. The results of investigations by several television networks now in progress should help to identify specific areas for further attention.

The broad areas of federal entitlement and grant programs, as well as government contracting programs, suggest numerous possibilities for the use of testing. Civil rights enforcement officials at several federal agencies have noted some interesting subjects for immediate exploration. For example, the pilot effort of testing nursing homes admission practices on the basis of race or HIV status undertaken jointly by the Office of Civil Rights at the Department of Human Services and the Civil Rights Division of the Justice Department could be expanded to include testing the admissions practices at government-supported hospitals and community clinics. The Department of Health and Human Services (HHS) might also use testing to examine possible discrimination in the administration of job referral and job placement programs under new welfare laws and regulations. HHS is already using testing to investigate the sale of cigarettes to minors.

The efforts now under way to use testing in connection with the Department of Agriculture's rural home loans program could stimulate further study of the uses of testing in connection with other DOA programs where discrimination has been alleged. DOA's program of loans for small farmers has already been the subject of litigation and seems appropriate for testing. Testing would also seem suitable to investigate possible abuses in the DOA's huge food stamp program and in connection with other Department loan programs.

Another area where testing might have special relevance is in government contracting. A substantial number of the factors identified as contributing to the underrepresentation of minority business enterprises in government contracts appear to be susceptible to testing. These include, but are not necessarily limited to, such practices as discrimination in the availability of bonding and credit, denials of product licenses or franchises, bid shopping, and double standards in the evaluation of performance. Not only could testing studies in any of these areas prove very valuable in clarifying issues around affirmative action, they might also produce strong evidence for civil rights enforcement action.

These suggestions are merely illustrative. It would be very interesting to survey civil rights enforcement officials throughout the federal government to determine the areas where they believe civil rights testing should be considered.

General Considerations for Future Civil Rights Testing

As testing gains greater acceptance as a civil rights research and enforcement technique, it is extremely important that it be used as effectively and as responsibly as possible. It is thus most appropriate that the primary entities now using testing in their work—HUD, DOJ, and the private fair housing councils affiliated with NFHA—have all stated their commitment to promoting high standards



for testing administration and evaluation. While some professional differences of opinion in this field are inevitable, the effective level of communication among key players is encouraging. Naturally, as the areas and quantities of testing expand, the need for coordination and communication in the field will become even greater.

The important employment testing initiatives supported by EEOC and OFCCP highlight the need for communication and coordination among all federal agencies that are using or considering the use of testing. They also reinforce the importance of strengthening the capacity of the FEC to serve as a national clearinghouse for private entities now using EEO testing as part of their work or considering the use of this technique in the future. It is also important to emphasize that government-supported testing is focused on research and compliance, rather than enforcement. For this reason it is probable that advances in EEO enforcement testing and support for establishing an effective monitoring clearinghouse and training capacity for private EEO enforcement work will require significant foundation support in the years immediately ahead. Support for private enforcement work and groups such as the FEC and NFHA is vital for another reason as well. The experience with testing to date confirms that the use of this approach in enforcement requires particularly careful planning and execution. To the fullest degree possible, it is important to build on the litigation successes achieved in housing and the initial employment cases in a coordinated manner. Communication among all interested parties is vital to this process.

As testing expands into new areas it is also important to consider what entities should be conducting the tests. The wealth of experience developed by private agencies in housing matters and in public accommodations testing strongly supports the idea that these entities should be a central part of this effort. State and local agencies also need to become more active in this field. The cooperative links that already exist between the groups doing EEO testing and the National Fair Housing Alliance and its affiliates will become even more important in the years ahead. It may be an appropriate time to consider formal working agreements between these groups and possibly mergers at the local level.

There are many reasons to be optimistic about the future of civil rights testing. The commitment to testing by HUD through its FHIP program and its strong endorsement by Secretary Cuomo are the main reasons to be encouraged. Similarly, the innovative work under way at the Justice Department is very positive. In addition, the increasing support for testing at EEOC and OFCCP has enormous potential. The policy discussions at HUD and at other federal agencies concerning the use of testers as part of a national audit of discrimination and civil rights enforcement further illustrates the recognition of how important testing can be, as is the attention testing is receiving as part of the President's Race Initiative. All of these developments reflect a growing recognition that, properly employed, civil rights testing can play a unique and essential part in educating the American public on the extent to which bias persists in our society and in enforcing our civil rights laws.



Endnotes

1. *Havens Realty Corp. v Coleman*, 445 U.S. 363 (1982).
2. NFHA lending testing formed the basis of complaints against four of the nation's largest mortgage companies. The insurance testing led to conciliated agreements with State Farm and Allstate, the nation's two largest homeowner insurance companies, requiring the companies to make major changes in their underwriting practices.
3. The HMDA, 12 U.S.C. Sec. 2801 et seq., was originally passed by Congress in 1975 and amended in 1989. The law requires lending institutions to report to federal bank regulatory agencies certain specified information about the home mortgage loans that they both make and deny. The information to be reported includes the race, national origin, sex, and income of the loan applicant, the amount of the loan, and census tract of the property.
4. Enforcement is conducted under Section 704(c) of the Equal Credit Opportunity Act (ECOA), 15 U.S.C. Sec. 1691c, as amended, in conjunction with Sections 5(m)(1)(A), 9, 13(b), and 16(a)(1) of the Federal Trade Commission Act (FTC Act), 15 U.S.C. Secs. 45(m)(1)(A), 49, 53(b), and 56(a)(1), as amended.
5. In particular, HUD officials would like to see the development of clear standards for administering tests in the form of "best practices," while NFHA is developing a certification program for its affiliates covering testing procedures and practices.
6. Over the past seven years the FEC has completed more than 2,000 tests and published three studies based on testing results involving race, age, and national origin.
7. *Fair Employment Council of Greater Washington, Inc. v BMC Marketing Corp.*, 28 F.3d 1268 (D.C. Cir. 1994); *Molovinsky v Fair Employment Council of Greater Washington, Inc.*, 683 A.2d 142 (D.C. Ct. App. 1996).
8. The EEOC had stated its general support for the use of testing in the form of a Policy Guidance issued in 1990 and updated in 1996.
9. *Floyd-Mayers v. American Cab Co.*, 732 F. Supp. 243 (D.D.C. 1990)
10. *U.S. v Flagstar Corp.* (Civ. No. 93-20208-JW) (N.D.Cal.), consolidated with *Ridgeway v Flagstar Corp.* (Civ. No. 93-20202-JW) (consent decree entered May 24, 1994); *Dyson v Flagstar Corp.* (C. A. No. DKC-93-1503) (D.Md.) (consent decree entered May 24, 1994).

References

- Ayres, Ian. 1991. "Fair Driving: Gender and Race Discrimination in Retail Car Negotiations." *Harvard Law Review* 10:817-72.
- Ridley, Stanley, James A. Bayton, and Janice Hamilton Outtz. "Taxi Service in the District of Columbia: Is It Influenced by Patrons' Record and Destination?" Report available from the Washington Lawyers' Committee for Civil Rights Under the Law. Mimeographed.



ABOUT THE EDITORS

Michael Fix is a lawyer and principal research associate at the Urban Institute, where he directs the Immigration Studies Program. His recent work has focused on discrimination measurement and immigration policy. He has written or edited five books on immigration, all published by the Urban Institute Press, including *Poverty amid Prosperity: Immigration and the Changing Face of Rural California* (1997) and *Immigration and Immigrants: Setting the Record Straight* (1994). He has co-authored or co-edited several titles on discrimination measurement: *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring* (1991); *Clear and Convincing Evidence: Testing for Discrimination in America* (1993); and *Do Minority-Owned Businesses Get a Fair Share of Government Contracts?* (1997).

Margery Austin Turner directs the Metropolitan Housing and Communities policy center at the Urban Institute, where her research focuses on spatial and racial dimensions of anti-poverty policies within urban regions. Ms. Turner served as Deputy Assistant Secretary for Research at the Department of Housing and Urban Development from 1993 through 1996, where she focused HUD's research agenda on the problems of racial discrimination, concentrated poverty, and economic opportunity in America's metropolitan areas. She co-authored a national housing discrimination study, which used paired testing to determine the incidence of discrimination against African-American and Hispanic home-seekers, and applied the paired testing methodology to employment, producing the first direct statistical evidence of discrimination against young African-American jobseekers. Ms. Turner also directed research on racial and ethnic steering and conducted an exploratory study of differential real estate marketing practices in minority and white neighborhoods.



ABOUT THE CONTRIBUTORS

Timothy Bates is professor of labor and urban affairs at Wayne State University in Detroit. He is a leading expert on minority business and on discrimination and other barriers to their formation and growth. He recently served as consultant to the United States Department of Justice in the development of new procurement regulations that respond to the Supreme Court's Adarand decision.

Marc Bendick, Jr., a labor economist, is a principal in Bendick and Egan Economic Consultants, Inc., in Washington, D.C. His research and policy analysis focuses on employment, poverty, and discrimination, applying testing to those issues for the past 10 years. Among his recent publications are *The Documentation and Evaluation of Anti-Discrimination Training in the United States* (International Labour Office, 1998) and *Connecting Inner-City Youth to the World of Work* (Committee for Economic Development, 1997). He can be reached at www.bendickegan.com.

Roderic V. O. Boggs is the executive director of the Washington Lawyers' Committee for Civil Rights. He has been actively involved in the development of civil rights enforcement testing for over 15 years as counsel to both the Fair Housing and Fair Employment Councils of Greater Washington. Among his publications is "Use of Testing in Civil Rights Enforcement" (with Joseph Sellers and Marc Bendick, Jr.) in *Clear and Convincing Evidence: Measurement of Discrimination in America* (Urban Institute Press, 1993).

Christopher F. Edley, Jr., is founding co-director of the Civil Rights Project at Harvard Law School. He specializes in research on administrative law and problems of governance, race, and public policy. His latest books include *Not All Black and White: Affirmative Action, Race, and American Values* (Hill and Wang, 1997), which grew out of his work as special counsel to President Clinton, and *Administrative Law: Rethinking Judicial Control of Bureaucracy*



(Yale University Press, 1990). He is a consultant to the President's Advisory Board on Racial Reconciliation.

Peter Siegelman is an economist who has taught at the Yale and University of Connecticut Law Schools and at Amherst College. He has conducted empirical research on employment discrimination litigation, discrimination in the sale of new cars, the economics of crime, and externalities in the provision of radio broadcasting.

John Yinger is professor of economics and public administration and associate director of the Center for Policy Research at the Maxwell School, Syracuse University. He is a specialist in racial and ethnic discrimination in housing, state and local public finance, and urban economics. His most recent book, *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*, won the Myers Center Award for the Study of Human Rights in North America.





THE URBAN
INSTITUTE