In Part 1 of this article two definitions of economic discrimination were given, along with a variety of statistical tables that illustrated them. Societal economic discrimination was defined as the difference in (or ratio of) the average family income of minority and majority groups. Income is the main component of economic well-being, and income tends to be positively correlated with other components, such as leisure, favorable nonpecuniary aspects of one's job, and so on. The statistical tables in Part 1 showed large income disparities between white and black families, between white non-Hispanic families and Hispanic families, and between families with a male primary earner compared to families with a female primary earner (or head of household). These disparities in income are widely perceived as inequitable.

Economic discrimination was also defined as the difference in (or ratio of) average wage rates of minority and majority workers who may be reasonably assumed to have equal productive capacities. This concept of economic discrimination has theoretical as well as practical importance because it challenges a fundamental principle of the workings of competitive economies: that equally productive workers should receive equal wages. It is essentially equivalent to the proposition that the same good or service (like labor) should receive the same price (or wage) in a competitive market.

The practical importance of wage, or labor market, discrimination is that wage rates or earnings—the latter being the product of the hourly wage and hours worked—are the most important component of income. Even such sources of nonlabor income as pensions, disability insurance, and unemployment compensation may be considered earnings-based, sometimes as deferred earnings or as an insurance payment derived from one's earnings. Earnings also reflect self-support and economic independence and therefore carry psychological benefits.

If equally productive workers are systematically paid unequal wages, there is a prima facie case for inefficiency, in addition to inequity, in the workings of the labor market. A condition may be said to be economically inefficient if the economy's output (or income) is less than it would be if that condition were eliminated, which implies that the costs of eliminating it are more than offset by the increased output that would result. As examples, inclement weather is not economically inefficient, whereas monopoly usually is. Whether labor market discrimination is economically inefficient depends on the theory and evidence one adopts. Economic efficiency need have no direct relation to ethical standards.

Persistent wage differences between workers of different ethnic and gender groups were shown in Part 1. Whether they reflect wage discrimination depends on how well productivity is measured and controlled for in the empirical evidence. The empirical research is complicated and controversial, in part because of data limitations and in part because of inadequacies in prevailing theories. This article summarizes the main theories and empirical research. The material is less well suited to simplification and summary, however, than were the descriptive statistics and definitions in Part 1.

Theories of labor market discrimination

The uses of theories

There are many economic theories of discrimination. They consist of models that specify selected features of labor markets in combination with selected theoretical principles drawn from a larger body of "textbook theory." The models are listed in Table 1 and will be discussed below. In this section on the uses of theories, references will be made to both these models and to principles from general economic theory.

We use theories to address the normative problems of inequity and inefficiency associated with labor market discrimination and to predict and explain the existence and persistence of discrimination. A theory should be testable with available data and point to policies and even remedies. Ideally, the quantifiable application of the theory should be translatable into measures of the costs and benefits of specific policies.

To illustrate, the theory of competitive markets predicts equal pay for equally productive workers. Assume that labor market discrimination is measured by the ratio of
minority wages to majority wages in a given labor market. Now consider testing the hypothesis that discrimination is greater when monopoly is greater, using data for different markets. A refined empirical investigation might quantify the relation between competition and discrimination so that we could estimate, say, the increase in the ratio of minority wages to majority wages as competition is increased. Ideally, the relation could be supplemented with additional information that reveals the costs of reducing monopoly and the benefits of the resulting increase in the ratio. (The costs and benefits do not have to be in dollars—perhaps “political capital” or votes are the relevant coin—but they do have to be in commensurable units if they are to be compared.) In principle, the benefit/cost calculation could be carried out for different strategies, and the most efficient strategy could be granted the highest priority in policy actions.

Unfortunately, the foregoing illustration is not easily applied. Current economic theories leave open the possibility of special cases in which competition is consistent with labor market discrimination, and the theories do not conclude that monopoly necessarily produces labor market discrimination. Furthermore, it is extraordinarily difficult to establish causal relations between variables like degree of competition and wage ratios with current methods of economic research.

Another illustration is the theory of wage determination, especially the part of this theory that involves human capital investments. This theory specifies a positive relation between wages and such assumed causes of productivity in the labor market as the worker’s education and training. Empirical investigations of these productivity/wage relations for minority and majority groups attempt to discover the sources of wage discrimination and to provide benefit/cost information to guide policies to remedy discrimination. This type of research has been abundant, but there is considerable controversy about how successful it has been, particularly in its policy implications.

Basic concepts of economic theories

Economic theories of discrimination deal almost exclusively with discrimination in the labor market and they deal almost exclusively with the demand side of the market. The theoretical challenge is to explain how workers who are intrinsically equal in productivity receive unequal wages. Thus, the supply side of the market is effectively neutralized by the assumption of either equal productivity or “controlled-for” productivity differences.

Discrimination in demand can be seen as a willingness-to-pay to avoid contact with the minority group or, equivalently for my purposes, a willingness-to-pay for contact with the majority group. This specification, which is due to Gary Becker, expresses and measures prejudice as a taste (preference) in money terms. This definition also involves the central principle by which discriminatory outcomes tend to diminish if competition in markets is assumed: namely, that many producers and mobility among economic agents will lead to the separation of groups to avoid the costs of contact. The separation does not imply autarky; trade between the groups continues. Once the groups are separated, economic discrimination disappears, given the definition of economic discrimination as different wages for equally productive workers. Thus, segregation is a mechanism for eliminating discrimination in competitive markets. This is a rather depressing conclusion for those who favor competitive markets and an integrated society as well as the elimination of discrimination.

But segregation is not the only way to eliminate discrimination in competitive markets. Collective action to offset the effects of discriminatory tastes or changes in those tastes can be accomplished without seriously restricting competition in markets. Indeed, common sense and casual observation indicate that an integrated society is generally more competitive. Nor is segregation necessarily a concomitant of no discrimination, as witness the Republic of South Africa.

Indeed, the conclusion about the incompatibility of discrimination with theories of competitive markets is not agreed to by all economists. There is not the space to defend the conclusion in any detail, but see the useful articles by Kenneth Arrow and Finis Welch. If competition is not assumed, then there are several additional theories of discrimination.

Types of theories

Table 1 provides a taxonomy of theories of labor market discrimination. As discussed below, none of the theories listed is widely accepted as a satisfactory explanation of the observed outcomes in the labor market. The problem is not that the theories lack logical consistency; rather, that none has convincing empirical support—a point reemphasized in the next section on empirical analyses. Consequently, the economist is not on firm ground when called upon to suggest policies for reducing the disparities in economic outcomes due to discrimination.

Neoclassical theories. The neoclassical theories of discrimination in competitive markets by consumers, workers, and employers imply that there will be no long-run (sustained) wage differential between equally productive majority and minority workers. Since competitive markets will reward the least-cost producer (indeed, the least-cost producers are the only survivors) and a necessary condition for least costs is that majority workers be paid no more than equally productive minority workers, it follows that wage differentials will disappear.

Using the case of consumer discrimination as an illustration, assume that all workers are equally productive and that consumers (who are predominantly white) are willing
Table 1

A Taxonomy of Theories of Labor Market Discrimination

Neoclassical Theories

A. Exact Models: Assumes perfect information

Competitive Theories
No monopolies or collusive behavior among economic agents. Sources of discriminatory preferences may be
1. Consumers
2. Workers
3. Employers

Monopoly Theories
Exclusive control by one person or group. Control may be exercised
1. By the firm over the product’s price (only one seller)
2. By employer over workers’ wages (monopsony, only one buyer)
3. By workers over wages (trade unions)
4. By government over a variety of market conditions (e.g., wage regulation)

B. Stochastic Models: Information lacking in some respect

Theory of Statistical Discrimination
In the absence of full knowledge of the workers’ productivity, firms rely on observable characteristics (race, sex, age) to estimate productivity.

Institutional Theories
Characterized by reliance on historical studies, legal analysis, or case studies.

Capable of describing combined forces of monopolies, discriminatory preferences, and particularistic circumstances, but no generalizable theory is generated.

A. The assumption of perfect information is equivalent to the assumption that the expected values (or means) of the variables fully describe the outcomes of interest.

b. Variability in the values of the variables, in addition to their means, may determine the outcomes, and the mean or variance may be unknown to the decision-makers.

say, clothing or automobiles according to the color of the workers in clothing or automobile factories. For these goods the price would simply be \( p \), regardless of the color of the workers. Therefore, black workers would specialize in the production of goods with no customer contact and, in so doing, avoid being paid a wage lower than that of an equally productive white worker, which would be the outcome if they competed with whites in, say, retail selling, where there is customer contact. Competition, activated by worker mobility and the incentives of firms to produce their product at the lowest cost, eliminates the discriminatory wage difference. This model does result, however, in segregation of the workers by industry.

Similar motivations, actions, and outcomes characterize the response to worker or employee discrimination. If white workers discriminate against black workers by acting as if they require a higher wage to work with black workers, then the labor force will become segregated by color, but there should be no wage differential by color for equally productive workers.

If employers discriminate against black workers by acting as if the labor cost (wage) of a black worker is higher than the labor cost (wage) of an equally productive white worker, then the wage rate of white workers will be higher than that of black workers. The money profits for employers hiring blacks will be higher, but this only “compensates” these employers for their distaste. However, any employer who does not discriminate will be able to undersell his competitors. If there are some nondiscriminatory employers, they become the only survivors, and the extinction of the discriminatory employers results in the disappearance of wage differences by color for equally productive workers.

As Arrow emphasizes and demonstrates, employer discrimination could result in a sustained wage difference only if no employers could be found who were nondiscriminatory.

Now consider the neoclassical monopoly models. Each offers the possibility for sustained discrimination, but none has persuasive empirical support.

Product monopoly does not imply monopoly power in the labor market. The monopolist must have the power to determine wages and must be willing to forgo money profits to “overpay” white workers (or male workers, etc.), and the monopolist must be willing to repel the efforts of nondiscriminating capitalists from taking over and increasing the monetary return on the investment. Surely the stockholders of a monopoly corporation desire to maximize profits. These considerations imply limited scope for discrimination due to product monopoly.

Monopsony, in which an employer is the sole buyer of labor in a market, is theoretically important, because it is the neoclassical model of exploitation. Workers are captive in a market where there is only one employer, or where a
group of employers collude and act as one buyer. Monopsony represents a rare area of common ground between neoclassical and Marxian models of the labor market. I doubt, however, that the monopsony model is empirically important in modern times, when markets are larger, the one-industry town has declined, and workers are more mobile than they were in decades past.

Workers' monopolies—trade unions—are potentially a source of discrimination against minority workers. We know that unions attempt to gain economic rents for their members in the form of above-competitive wages, and that this requires that the unions must limit entry. Thus, the union’s control over entry, its domination by majority-group workers, and its ability to raise wages above competitive levels give the majority group the capacity to discriminate against minorities without being at a competitive disadvantage. Historically, American trade unions have been guilty of many specific acts of discrimination against minority groups. Yet the most thorough empirical study of the effects of unions on white-black and male-female wage differences does not show that unions are an important source of economic discrimination. This study will be discussed in the next section.

The government may regulate labor markets in ways that promote or retard the status of minority workers. A much-discussed example is the minimum wage law, which may prevent the competitive principle of least-cost production from operating. In particular, the law could prevent a discriminatory advantage to majority workers from being eliminated by competition. On the other hand, governments have enforced laws against discrimination and have aided minority workers in other ways. The net result of these conflicting policies is not clear.

In conclusion, there is surprisingly little empirical evidence that the various monopoly theories explain much of the labor market discrimination that exists and has persisted in the United States.

My first category of neoclassical models were those constructed on the basis of complete information. In practice, the variables cannot be known with certainty. The theory of statistical discrimination is based on this uncertainty and, therefore, has an initial appeal. Because firms must hire, pay, and promote workers without perfect knowledge about the workers' productivity, employers rely on observable demographic characteristics as indicators of productivity. Thus, if employers believe black workers are, other things equal (such as, say, years of schooling, etc.), less productive than white workers, they will pay blacks less. Moreover, various conventional indicators of productivity may be less reliable for minority workers than for majority workers. Despite these conditions, the two postulates of the competitive model—large numbers of firms and the survivor principle for least-cost producers—will lead to a tendency for average payments to workers to equal their average productivity. The statistical uncertainties affect the groups' variances (or dispersion) of wages, but not their averages.

Institutional theories. Institutional theories of discrimination are a varied group of historical, legal, and case-study analyses of labor market discrimination. They lack a formal structure and are limited in their generalization. At the same time these studies are able to deal with more complicated structures than the economic neoclassical models; they may describe the interrelations of the combined forces of, say, monopolistic industries, trade unions, government regulation, and community prejudices. I believe that there are many useful and persuasive examples of discrimination in the institutional literature.

Empirical analysis of labor market discrimination

Aside from simple descriptive statistics, empirical research on labor market discrimination may be divided into (1) tests of hypotheses suggested by the theories, such as the proposition that discrimination is less in competitive industries, and (2) estimation of the amount and determinants of discrimination; for example, estimating the change in the relative wages of minority workers over time, over the course of the business cycle, or in different industries, and so on.

Testing of hypotheses

Hypothesis testing has been, as Masters noted, “surprisingly limited” and has produced few firm conclusions. One reason is that the theories often yield ambiguous predictions. Discrimination may, for example, be predicted to exist in the short run but not in the long run, with no basis for determining the time required for the transition. Also, the theories suggest many economic influences, and the empirical work usually concentrates on one influence in isolation.

Ashenfelter's previously mentioned study is one of the few that test hypotheses with convincing results. Ashenfelter tested the hypothesis that labor unions are a source of economic discrimination. He found that the white-black wage difference was reduced, rather than increased, by unions, because black workers are somewhat more likely than whites to be union members, and the effect of unions on wages is somewhat larger for blacks than whites. On the other hand, unions were found to increase slightly the wage difference between men and women, because women are less likely to be union members and, when they are union members, their wage gains are smaller than those of men.

The importance of Ashenfelter's study is that it offered no support for a neoclassical hypothesis of discrimination by unions, which appeared stronger theoretically than the hypotheses rationalizing discrimination by other economic agents—consumers, employers, or governments.
Estimation of discrimination

In Part 1 of this article I mentioned that wage discrimination against a group was measured empirically by a “negative” effect of group status on wages, after controlling for productivity. Typically, a statistical regression function is used to estimate the effect of group status on wages, and the control over productivity, as measured by various characteristics of the workers, is handled by this statistical technique. Such a technique attempts to measure a remaining or residual difference between minority and majority workers’ earnings by equalizing (holding constant) the factors that determine a person’s productivity. Let us call this residual difference a difference in predicted earnings. It turns out that the ratio of minority workers’ predicted earnings to majority workers’ predicted earnings varies widely, partly reflecting variation in the data sources, and more important, partly reflecting the theoretical ambiguity about the proper set of variables to hold constant in attempting to equalize productivity. Furthermore, the estimated difference in predicted earnings can be biased as a measure of discrimination because some of the factors that determine productivity (such as training) may in themselves reflect discrimination. A second type of bias occurs because inevitably some factors that determine wages are not taken into account. Obviously a statistical technique is restricted to only those variables that can be observed and measured. There are bound to be omitted variables, some of which may be known to the worker and employer, but not to the statistical analyst.

Sometimes the statistical analyst will merely assume that the omitted productivity variables lead to a bias such that labor market discrimination is overstated. The analyst may assume, for example, that men are more productive than women in ways not measured by the variables in the statistical model predicting earnings. Clearly, if this is assumed, it follows that labor market discrimination has been overstated. However, I do not believe that the omitted variable problem should be referred to as a systematic bias. The omission of variables can lead to a bias in either direction.

Marketwide studies. A number of marketwide studies of discrimination have been carried out. These studies usually report the observed, unadjusted ratios of minority workers’ wages to those of majority workers—the ratio that is measured without adjusting for any productivity variables. This ratio will typically be around .6, as was shown in Part 1 for the earnings ratios of women to men and of black men to white men. This ratio rises to .7 when exogenous variables (those not affected by discrimination) such as age, years since immigration, region of residence, and so on, are held constant. The ratio rises again, to .8 or .9, when such variables as industry, occupation, and years with the firm—which in my view are endogenous to the process of discrimination under analysis—are held constant.

Several analysts claim that a ratio of around .9 is found for Hispanic wages compared to non-Hispanic white wages, holding constant the following variables: age (Hispanics tend to be younger); education (Hispanics tend to have less schooling); years of residency in the United States; and a variable measuring whether English was the primary language spoken at home when growing up.

When comparing women and men, the ratios rise from .6 to .8 or so when variables like marital status, numbers and ages of children, hours worked per year, and years of labor market experience are held constant. Note that each of these variables might be considered to reflect labor market discrimination—that is, women work less in the market and more at home because they are not offered employment opportunities and wages equal to those of men. Note also that controlling for age and education would not much affect the ratios, because the means of these variables tend to be the same for men and women.

Studies of individual firms. The statistical model is also used to analyze discrimination in individual firms. In the United States such analyses are sometimes offered as evidence in litigation stemming from antidiscrimination laws.

Again, the model holds constant those worker characteristics that are assumed to represent productivity—here, productivity to the firm. As noted in Part 1, many variables, like years of schooling, which may reflect marketwide discrimination, are clearly exogenous to the individual firm. Moreover, if the issue is “fairness” in the treatment of employees, rather than fidelity to an abstract ideal of “true productivity,” then the selection of variables may be determined from the employer’s explicit criteria for hiring, retention, promotion, and pay. These criteria can be specified with relative precision. They may be examined to determine if they do or do not reflect employer discrimination.

Unfortunately, the analyses of data from a single firm have two serious faults that limit their use for assessing marketwide discrimination. First, the sample is small and nonrandomly selected. Data for one company refer to only one industry and a few occupations, and the role of market discrimination in determining the allocation or distribution of minorities among industries and occupations is not examined. Second, we seldom know the selection rules that determine how the workers become applicants to or attached to the firm. Nor do we know whether the company’s tactics of, or reputation for, discrimination affect the number and composition of minority workers who apply for jobs at the firm.

The value of empirical analysis

I conclude that the estimation procedures discussed in this section serve the following purposes: (a) they provide a way of monitoring discrimination over time and in different
contexts; (b) they may suggest policy variables to manipulate by showing which productivity characteristics have a large effect on earnings; (c) they help to determine whether an individual firm is discriminating. Nevertheless, the empirical research taken as a whole does not have a solid theoretical foundation, and the research requires subjective interpretations.

Welfare implications and conclusions

Variation in the analyses of different groups

Inequities in economic well-being among racial, ethnic, and gender groups appear to be widespread, and economic theories for why they persist are only moderately helpful. Empirical research on the discrimination experienced by different groups has yielded explanations of varying levels of satisfaction.

The case of blacks in the United States offers the strongest evidence for labor market discrimination and, given existing theories, for flaws in the competitive functioning of the market. By contrast, the discrimination believed to have faced such immigrant nationality or religious groups as Irish Catholics, Italians, Japanese, and Jews some 50 to 100 years ago seems today to have been overcome with respect to income and earnings. This evolution toward equality with whites whose ancestry was Anglo-Saxon is consistent with a neoclassical view of the workings of competitive markets, assuming that the productive capacities of the different ethnic groups are equal and that the economy is sufficiently competitive.

The disparities in wages between men and women may be rationalized by the argument that specialization in work in the market sector rather than the home sector leads to higher market wages for men. However, this hypothesis must rely on unobserved productivity characteristics, because when observed characteristics are held constant, a wage disparity between men and women remains. Furthermore, the argument about specialization does not explain why the total income received by women during their lifetimes is less than that for men, as was shown in Part 1.

The lower earnings of Hispanic Americans relative to white non-Hispanics may be explained by the importance, in determining earnings, of information about the labor market, facility in the English language, and years of schooling. The theories postulating these determinants of earnings for Hispanics and white non-Hispanics are qualitatively supported by empirical evidence, but I doubt that the quantitative gap in earnings is well explained by these theories.  

The wages, earnings, and incomes of black workers and black households are substantially less than those of whites, and the conventional human capital variables, such as education, training, and health care, do not explain much of the difference. Even if they did, the question would then be, Why is the market for such human capital investments functioning so poorly that blacks continue to be shortchanged? If whites find that these investments in human capital result in higher earnings and better jobs, why are blacks' opportunities for these investments so curtailed? If the answer is not labor market discrimination, is it discrimination in the capital markets that supply funds or sources of human capital investments? It is not scientifically satisfactory for economists to argue that labor market discrimination is minimal, if they have no explanation for how discrimination in capital markets creates and sustains the disparities we measure in the labor market.

The effect of discrimination on total output

One issue that has not been much studied is the implications of discrimination for economic efficiency, as measured by the size of total societal income.  

The neoclassical economist's convention (perhaps it is an obligation) to take tastes—individual preferences—as given, virtually prevents the translation of "different prices (wages) for the same good (labor)" into a loss in total societal income, or deadweight loss. Thus, there is no presumptive case for inefficiency in a competitive economy in which tastes are the fundamental cause of discrimination. Surely something is amiss. Discrimination in its many forms, not only economic, is widely believed to suppress the achievements of the minority group with no fully offsetting gains to the majority group. The economists that I know agree with this belief, yet conventional economic theories do not, to my knowledge, explain or analyze this widely shared conviction.

Economists have prescribed limits for themselves in many policy spheres. Economics does not distinguish among the ethical merits of different tastes; between, say preferences for physical attractiveness or for race. As economists we have nothing to say about the justness of laws that prohibit an employer from refusing to hire someone on the basis of color but that permit hiring on the basis of physical attractiveness. As citizens we may, of course, have strong opinions about such matters.

Instead, the role of economic analysis lies in the measurements and methods that permit prediction. Empirical regularities such as time trends may be useful even in the absence of fully developed theories. At a minimum, the measurements provide valuable theories. A more ambitious form of empirical research is that aimed at evaluating government policies that attempt to reduce discrimination and to offset its outcomes. The essential difficulty in evaluating these programs is the classic problem of making inferences from an uncontrolled experiment. We observe an outcome for a group of workers, some of whom participated in the program or, alternatively, had the program imposed on them. To establish causality between
program status and the outcome, the factors that selected the workers into the program must be either (a) known and controlled for in the evaluation, or (b) known to be unrelated to the outcome.

It is difficult to know enough about the selection process and about all the causes of the outcome to satisfy either condition (a) or (b). Random assignment would satisfy condition (b), but this selection procedure is rare. Legislators and courts, therefore, seldom rely on the research of economists to determine the fate of government programs.

Final word

The economics of discrimination is a particularly complex subject. Theories of discrimination have been useful for providing definitions and for suggesting measurements of discrimination but not for providing convincing explanations of the phenomenon or of its patterns. The econometric work has also been useful, but more for its descriptive content than for testing hypotheses or for providing estimates of causal relations.

1A longer discussion of these topics is given in Cain, "The Economic Analysis of Labor Market Discrimination: A Survey" (see box).
5These ideas, along with some special cases of sustained discrimination, are presented by Dennis J. Aigner and Glen G. Cain in "Statistical Theories of Discrimination in the Labor Market," Industrial and Labor Relations Review, 30 (January 1977), 175-87. As discussed in this article, the special cases in which the model predicts discrimination do not seem to be empirically important.
9Ibid.
10Ibid., see "Case 2."
11A large number of studies presenting empirical estimates of wage discrimination against women and blacks are summarized in Cain, "The Economic Analysis of Labor Market Discrimination."
14For a discussion of the empirical evidence about wage discrimination against Hispanics, see Cain, "The Economic Analysis of Labor Market Discrimination."
15I have elsewhere addressed this question regarding discrimination against women, and my conclusions were embarrassingly thin. See "Welfare Economics of Policies toward Women," Journal of Labor Economics, 2 (October 1984). Also available as an IRP Discussion Paper (see box).

ASPE-Institute workshop

Research on the Labor Market and Program Participation of Hispanics, Immigrants, and Southeast Asian Refugees. This project examines the income and employment experiences of these groups and compares them to the native population.

• Robert Bach (State University of New York at Binghamton) and Marta Tienda (IRP), "Contemporary Immigration and Refugee Movements and Employment Adjustment Policies."
• George Borjas (University of California-Santa Barbara), "The Impact of Assimilation on the Earnings of Immigrants: A Reexamination of the Evidence."

An additional paper was contributed by Robert Lampman (IRP), "Some Topics in Need of Poverty Research: Respect and Prospects."

The discussants included Irwin Garfinkel (IRP), Paul Menchik (Michigan State University), Maurice MacDonald (IRP), Christopher Flinn (IRP), Marcia Weaver (U.S. Department of Health and Human Services), Robert Moffett (Brown University), Gary Sandefur (IRP), Bernard Stumbras (Wisconsin Department of Health and Social Services), George Borjas, and Daniel Weinberg (U.S. Department of Health and Human Services).