in Orley Ashenfelter and Albert Rees, eds.
Discrimination in Labor Markets, 1973

Kenneth J. Arrow

THE THEORY OF DISCRIMINATION

1. INTRODUCTION

The fact that different groups of workers, be they skilled or unskilled, black or white, male or female, receive different wages, invites the explanation that the different groups must differ according to some characteristic valued on the market. In standard economic theory, we think first of differences in productivity. The notion of discrimination involves the additional concept that personal characteristics of the worker unrelated to productivity are also valued on the market. Such personal characteristics as race, ethnic background, and sex have been frequently adduced in this context.

Discrimination in this paper is considered only as it appears on the market. Obviously, one can have discrimination in the same sense whenever decisions are made that concern other individuals, namely, when personal characteristics other than those properly relevant enter into the decision. Deliberate racial segregation and discrimination in entrance to schools and colleges, deprivation of the right to vote along social and sexual lines, and discriminatory taxation are all examples of nonmarket discrimination.

It may as well be admitted that the term "discrimination" has value implications that can never be completely eradicated, though they can be sterilized for specific empirical and descriptive analyses. I have spoken of personal characteristics that are "unrelated to productivity" and not "properly relevant." These terms imply definitions of product and of relevancy which are themselves value judgments or at

any rate decisions by the scholar. The black steel worker may be thought of as producing blackness as well as steel, both evaluated in the market. We are singling out the former as a special subject for analysis because somehow we think it appropriate for the steel industry to produce steel and not for it to produce a black or white work force.

However, the value judgments are intrinsic only in determining which wage differences we regard as worth studying as an example of discrimination, *not* in the empirical or theoretical analysis of any form of discrimination once specified.

In the following, I will address myself specifically to racial discrimination in the labor market. For the most part, the analysis extends with no difficulty to sexual discrimination. The other markets in which discrimination has been most observed, especially housing but also insurance and capital, are analyzed by the same general methods, but the operation of these markets has led more often to simple exclusion and less to price differentials.

The basic aim here is to use as far as possible neoclassical tools in the analysis of discrimination. As will be seen, even though the basic neoclassical assumptions of utility and profit-maximization are always retained, many of the usual assumptions will be relaxed at one point or another: convexity of indifference surfaces, costless adjustment, perfect information, perfect capital markets. As I will try to show, the abandonment of each of these assumptions is motivated by a clearly compelling reason in the theoretical structure of the subject. Personally, I believe there are many other economic phenomena whose explanation entails the abandonment of each of these assumptions, so the steps proposed here are not ad hoc analyses but should be important elements in a more general theory capable of analyzing the effects of social factors on economic behavior without either lumping them into an uninformative category of "imperfections" or jumping to a precipitate rejection of neoclassical theory with all its analytic power.

The first application of neoclassical theory to discrimination that I know of is that of Edgeworth, but the main study to date has been that of Becker.[1] The analysis to be presented here appears in a more technical form in an earlier paper.[2] It seeks to develop further Becker's models and to relate them more closely to the theory of general competitive equilibrium, though frequently by way of contrast rather than agreement.

Since I am presenting here the theory of discrimination in the labor market and not the entire theory of racial differences in income, I abstract from differences in productivity between the groups of workers. In an empirical study, it will be necessary to allow for this possibility. In the case of blacks and whites, some possible causes of productivity differences have been established (differences in educational quantity and quality, family size and socio-economic status, and household headed by woman); and others surmised (culturally varying attitudes toward work and future-orientation derived from the heritage of slavery and other historical factors).[3] These differences themselves may

[1] See for example, F. Y. Edgeworth, "Equal Pay to Men and Women for Equal Work," *Economic Journal*, 31 (1922), 431-457; and Gary Becker, *The Economics of Discrimination* (Chicago: University of Chicago Press, 1959).

[2] For further analysis see Kenneth J. Arrow, "Models of Job Discrimination," Chapter 2 in A. H. Pascal (ed.) *Racial Discrimination in Economic Life* (Lexington, Mass.: D. C. Heath, 1972) pp. 83-102; and Arrow, "Some Models of Race in the Labor Market," Chapter 6 in A. H. Pascal, *ibid.*

[3] See for example, O. D. Duncan, "Inheritance of Poverty or Inheritance of Race?" in D. P. Moynihan (ed.) *On Understanding Poverty* (New York: Basic Books, 1969), Ch. 4, pp. 85-110. Although my concern here is with discrimination and not with productivity differences, I must note my skepticism about the frequently made argument that blacks have less future-orientation. For this disregards the well-known fact that at any given income level blacks save at least as much as whites. This remains essentially true even when "income" is understood to mean "permanent income"; see Milton Friedman, *The Theory of the Consumption Function* (Princeton, N.J.: Princeton University Press, 1957), pp. 79-85; and H. W. Mooney and L. R. Klein,

be the result of discrimination in other areas of life. But for theoretical analysis of discrimination in the labor market, it is legitimate to assume that there are two groups of workers, to be denoted by B and W, which are perfect substitutes in production.

For the simplest model, then, we have a large number of firms all producing the same product with the same production function. Discrimination means that some economic agent has some negative valuation for B or positive valuation for W, or both, a valuation for which the agent both is willing to pay and has the opportunity to pay. The agents who could possibly discriminate are the employer, who might sacrifice profits to reduce or eliminate B employment in his plant, or the W workers who might accept a lower wage to work in a plant with more W and less B workers. (It is also possible that, for products sold on a face-to-face basis, customers might discriminate by being willing to pay higher prices to buy from whites; this case could be studied along similar lines but will not be dealt with here.) Not all discriminatory feelings can find expression in the market; an entrepreneur who has a distaste for competing against firms with B workers has no way, within the economic system at least, of expressing his tastes and therefore of influencing wage levels.

I assume that, given the tastes, the markets work smoothly. General equilibrium requires full employment of both B and W workers; the wages of both will adjust to clear the market, and the discriminatory tastes will be reflected in wage differences.

Let us first consider the simplest case, that in which the employer discriminates. Then he accepts a trade-off between profits, π, and the numbers of B and W employees. That is, we suppose he seeks to maximize, not profits, but a utility function, $U(\pi, B, W)$. We assume, to get the sim-

---

"Negro-white Savings Differentials and the Consumption Function Problem," *Econometrica*, 21 (1953), 425-456.

plest case, that there is only one type of labor; in the short run, we also take capital as given, so that output is $f(W+B)$, since the two kinds of labor are perfect substitutes (at a one-to-one ratio). If we take output as numeraire, then profits are given by the expression

$$\pi = f(W+B) - w_W W - w_B B, \tag{1}$$

where w_W and w_B are the wage rates, taken as given by the employers. If we proceed along conventional lines, the employer equates the marginal productivity of each hand of labor to the price to him. But here the "price" of B labor is the market price, w_B, *plus* the price the employer is willing to pay, in terms of profits, for reducing his B labor force by one. This second term is what Becker has termed the "discrimination coefficient," to be designated as d_B; it is the negative of the marginal rate of substitution of profits for B labor. If, as we usually suppose, the marginal utility of B labor is negative, then the discrimination coefficient, d_B, is positive.

In symbols,

$$MP_B = w_B + d_B, \tag{2}$$

where $d_B = -MR_{\pi,B}$. Similarly,

$$MP_W = w_W + d_W, \tag{3}$$

where d_W is negative (or zero if the employer has no positive liking for having W workers). But we are assuming that the two types of labor are interchangeable in production, so that $MP_W = MP_B = MP_L$, say. Then, from (2) and (3), $w_W + d_{dW} = w_B + d_B$, or

$$w_W - w_B = d_B - d_W > 0, \tag{4}$$

so that equilibrium requires that W wages exceed B wages, as might be expected.

For the moment, assume that all firms have the same utility function, $U(\pi, B, W,)$. It then appears reasonable to assume that all hire the same amounts of B and W (but we

will return to this point in the next section). Then each firm's labor force is the same, and the allocation of labor is efficient. The effects of discrimination are purely distributive. The most obvious implication then is that B workers are paid less than their marginal product, so that the W workers and employers together gain. Also, the W workers clearly gain, or at least do not lose, from (3), with $d_W \leqq 0$. The effect on profits, however, depends on the exact nature of the utility function. Under the assumption made, it follows from (1-3) and the fact that $MP_W = MP_B = MP_L$ that,

$$\pi = f(L) - (MP_L)L + d_W W + d_B B, \tag{5}$$

where $L = W+B$, the total labor force of the firm. If there were no discrimination, profits would be,

$$\pi_0 = f(L) - (MP_L)L,$$

and therefore the change in profits is simply,

$$\pi - \pi_0 = d_W W + d_B B. \tag{6}$$

The right-hand term has a simple interpretation. If we consider an increase in the firm's labor force with the proportions of W and B workers constant, then the negative of the marginal rate of substitution of profits for this balanced increase is simply $d_W\ (W/L) + d_B\ (B/L)$; this is the firm's need for additional profits to compensate it for a balanced increase in size. This term may of course be positive or negative.

However, a plausible hypothesis which we shall maintain hereafter is that employers' satisfactions depend only on the ratio of B to W workers. In that case,

$$d_W W + d_B B = 0, \tag{7}$$

and (6) tells us that employers neither gain nor lose by their discriminatory behavior. The entire effect is that of a transfer from B to W workers.

Let us now relax the assumption that utility functions are identical among firms. We continue to assume that for each

firm, the utility depends only on the ratio of W to B workers, but some firms may be more discriminatory than others, in the sense that the marginal rate of substitution of profits for B workers will be more negative at any given ratio, B/W. Equations (4) and (7) hold for each firm, at least each firm that employs both types of workers. They can be regarded as a pair of linear equations in d_W and d_B, to yield,

$$d_B = W(w_W - w_B)/(W + B),$$
$$d_W = -B(w_W - w_B)/(W + B),$$

which can be rewritten,

$$W/L = d_B/(w_W - w_B);$$
$$B/L = -d_W/(w_W - w_B).$$

Since $d_B > 0$, if there are both B and W workers, it must be that $w_W > w_B$, as before. We will observe firms with different ratios of W to L. The firms that display the most discrimination at the margin, i.e. the highest values of d_B, have the highest ratios of W to L. Thus an observation on all the firms in existence at equilibrium will reveal a dispersion of W-proportions in the labor force, and these ratios will measure the varying degrees of discrimination. Thus a partial degree of segregation appears; the B workers tend to be found in the less discriminatory firms, the W workers in the more discriminatory ones.

However, further analysis leads to implications which might raise some empirical questions. Specifically, equation (2) still holds, with $MP_B = MP_L$. Hence, according to the model, MP_L is higher for more discriminating firms. But then if we assume diminishing marginal productivity of labor, it follows that, the less discriminatory the firm, the larger it will be. This accords with common sense; discrimination is costly to the entrepreneur and acts as a tax on him, since it shifts his demand for labor to the more costly component. Hence, it restricts his scale.

Since MP_L is no longer the same from firm to firm, it follows that production is no longer efficient. The previous

strong statements about the incidence of discrimination no longer hold exactly either. However, their general thrust is still probably correct. Efficiency losses are not apt to be great, and the main redistribution is still likely to be from *B* workers to *W* workers.

It has been seen that competition tends to reduce the degree of discrimination in the market, in the sense that the unweighted average of discrimination coefficients of the different firms exceeds the average weighted in proportion to the number of workers.

This result, which may or may not be empirically reasonable, appears more strongly and less likely when one pushes the analysis into the long run. Now we are assuming that capital, which has been hitherto held fixed, is adjusted optimally to the size of the labor force. Then capital will flow to the more profitable enterprises which, in this context, are the less discriminatory. In the long run, output is therefore simply proportional to labor (assuming the production function displays constant returns to capital and labor). The marginal product of labor is then constant. As a result, the competitive effect just studied assumes an exaggerated form. Only the least discriminatory firms survive. Indeed, if there were any firms which did not discriminate at all, these would be the only ones to survive the competitive struggle. Since in fact racial discrimination has survived for a long time, we must assume that the model just presented must have some limitation to which we will return in Section 4.

We have dealt extensively with the assumption of discrimination by employers. But, as we observed earlier, discrimination by co-workers is also a possibility. The most straightforward extension of the preceding analysis is to the case of complementary services. Suppose now there are two kinds of workers, say foremen and floor workers. It is the foremen who like working with *W*'s and dislike working with *B*'s. As before, we assume that the likes or dislikes are governed by the ratio of *W* to *B* floor workers. Each fore-

man then chooses among alternative employment opportunities on the basis of both wages and the W/B ratio. Assume that all foremen have the same utility function.

The equilibrium in this model is a trifle unorthodox. Instead of an equilibrium wage for foremen, there is an equilibrium relation between foremen's wages and W/B ratios in firms. Every firm must lie on this curve, and the equilibrium curve will be one of the foremen's indifference curves between wages and W/B.

Let F be the number of foremen, and w_F their wage. Then the firm faces fixed w_W and w_B for the floor workers and a fixed *relation*,

$$w_F = w_F(W/L) \tag{8}$$

where $L \equiv W+B$ is the total floor force. The firm's short-run profits are defined by,

$$\pi = f(L, F) - w_W W - w_B B - w_F F, \tag{9}$$

where it is assumed, as before, that W and B floor workers are perfect substitutes.

Assume now that firms have no discriminatory tastes. They seek only to maximize profits. They will still not hire B workers at equal wages with W since an increase in W decreases the wages and therefore the cost of F, while an increase in B increases the cost of F. Hence, a W worker is worth more than his marginal product, while a B worker is worth less, exactly as in the case of employer discrimination. Further, the extent of the premiums over or deficits from marginal product depends only on the ratio of W to B. Hence, the previous analysis applies with suitable modifications. W workers are paid more than their marginal product, B workers less. If all firms wind up with the same levels of W and B, then the results are entirely parallel to those for employer discrimination: production remains efficient, and the entire incidence of the foremen's discrimination falls negatively on the B workers and positively to an equal extent on the W workers.

As in the case of employer discrimination, the extent of the wage difference between B and W workers depends on the extent of discrimination. The precise formula is of some interest. Recall that, by (8), w_F is a function of the ratio, W/L. By w'_F, I will mean the derivative of w_F with respect to this ratio (this is negative). Then w'_F/w_F is the proportional rate of change of the demanded wage rate (along the equilibrium indifference curve between foremen's wages and W proportion in the floor force) and therefore is a measure of discriminatory tastes. Let S_F be total payments to foremen, S_L total payments to floor workers. Then the following has been shown:[4]

$$\frac{w_W - w_B}{MP_L} = -\frac{w'_F}{w_F}\frac{S_F}{S_L} \tag{10}$$

The left-hand side is the market wage differential due to discriminatory tastes of foremen relative to the wage level in the absence of discrimination.

This formula has an interesting aspect. Given the degree of discrimination as measured by $-w'_F/w_F$, the observed wage differential depends on the ratio S_F/S_L. That is, the more important the share of foremen in the output of the firm relative to floor laborers, the greater the wage differential.

The language of the preceding analysis has assumed that it is the foremen or other supervisory employees who discriminate according to the composition of the floor workers. But the analysis itself is completely abstract. It may be illuminating to reverse the roles. Suppose that production workers have strong discriminatory feelings about their supervisors. Certainly the idea that white workers strongly resent being bossed by black supervisors or male workers by female foremen (foreladies? forepersons?) is a common one. Then if in (10) we understand by W and B those kinds of supervisory workers, by L the total number of such

[4] Arrow, "Some Models of Race in the Labor Market," Section B.

workers, and by F the floor workers, we have an excellent explanation of discrimination against B supervisory workers, for S_F then would be very large indeed compared with S_L.

Foremen may possibly differ in their tastes for discrimination. One might suppose that this will lead to a reduction in market wage differentials, analogous to the situation with employer discrimination. But a fuller analysis of this case remains to be done.

2. NONCONVEXITIES IN INDIFFERENCE SURFACES AND OPPORTUNITIES

I have gradually become convinced that the usual assumption that indifference surfaces are convex is inapplicable to the case of racial discrimination and indeed to many other problems in the economics of externalities. Pollution provides another example; Starrett has already pointed to the importance of nonconvexity in this context.[5] Assumptions which seem very reasonable in the contexts of discriminatory behavior *necessarily* imply a nonconvexity of the indifference surfaces of the firms in the case of employer discrimination or of the firm's profit function in the case of discrimination by complementary workers.

Actually, my view is that nonconvexity of indifference surfaces is in fact a widespread phenomenon. An excellent example in commodities with no externalities is residential location. One *could* after all live half the time in one place and half in the other. Convexity implies that such an arrangement would be at least as good as the least preferred of the two locations. If one is indifferent to the two, then one will prefer the mixture. In fact, taken literally, convexity would imply that individuals would be willing to spend half of any minute in one place and half in the other.

[5] See D. Starrett, "Fundamental Non-convexities in the Theory of Externalities," *Journal of Economic Theory*, 4 (1972), 180-199.

But (except for a few "beautiful people") most individuals find it preferable to live in one place, even though there may be another to which they are indifferent.

Indeed, if one looks through the literature, it is hard to find a convincing intuitive explanation of convexity of indifference surfaces. The best argument is that convexity is a necessary and sufficient condition for the continuity of demand functions. But this argument applies only to individual demand functions. Since each individual is small on the scale of the entire market, even the largest discontinuity in an individual demand function implies a negligible discontinuity in the market demand function. Hence, observations which suggest approximate continuity in market demand functions in no way imply convexity of indifference surfaces. In particular, the existence of general competitive equilibrium remains unaffected, or, to be precise, the existence of an approximate equilibrium of supply and demand on all markets can be demonstrated. (This line of argument was suggested initially by Farrell and subsequently developed by Bator, Rothenberg, Aumann, and Starr; for one exposition, see Arrow and Hahn.)[6]

It is true that the market demand function, if it is effectively continuous, can be derived by adding up a new set of individual demand functions, each derived from a "convexified" indifference map obtained from the original by filling in all the holes in the indifference surfaces. From the point of view of prices and total market quantities, the new-

[6] For further analysis, see M. J. Farrell, "The Convexity Assumption in the Theory of Competitive Markets," *Journal of Political Economy*, 67 (1969), 377-379; Francis Bator, "Convexity, Efficiency, and Markets," *Journal of Political Economy*, 69 (1961), 480-483; Jerome Rothenberg, "Non-convexity, Aggregation, and Pareto Optimality," *Journal of Political Economy*, 68 (1960), 435-468; R. J. Aumann, "Existence of Competitive Equilibria in Markets with a Continuum of Traders," *Econometrica*, 34 (1966), 1-17; R. Starr, "Quasi-equilibria in Markets with Nonconvex Preferences," *Econometrica*, 37 (1969), 25-38; and Kenneth J. Arrow and Frank Hahn, *General Competitive Analysis* (San Francisco: Holden-Day, 1971), Ch. 7.

ly formed indifference map predicts as well as does the original, and therefore one might be tempted to assume that one could act "as if" indifference surfaces were convex, though with some flat surfaces. But there is a loss of information, for the *distribution* of goods among individuals is quite different from what it would be if all individuals had convex indifference surfaces. Thus, in our residential location example, the market totals (how many people-hours are spent in each place) and the rents in the two places are well predicted by the convex approximation. But recognizing the underlying nonconvexities enables us to predict that half the people will be in one place all the time and half in the other, instead of each individual's spending half his time in one place and half in the other.

Let me give a brief diagrammatic illustration. Suppose every individual has the same indifference map, as given by Figure 1, and the same initial endowment, represented by A. One's initial reaction, conditioned by years of working

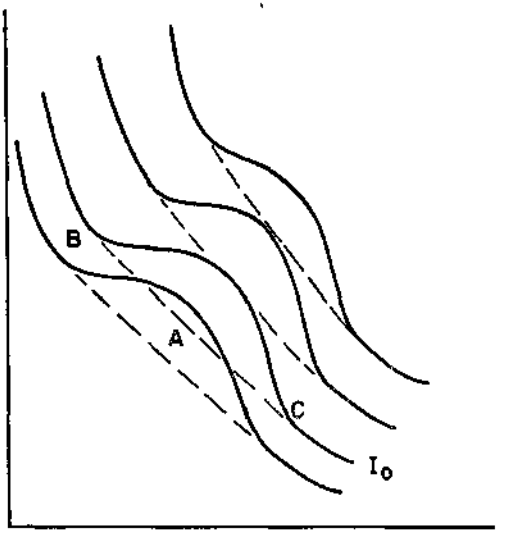

FIGURE 1

with convex indifference maps, is to assume that there is no trade; since all individuals are alike in every economic respect, they should wind up alike, which in this case means each with his own initial bundle. But this is clearly false. In fact the equilibrium can be obtained as follows: convexify each indifference curve by filling in the hole with a straight line segment tangent to the curve at both ends, as, for example, the segment BC on curve I_0. Now we see that, if we pretend for the moment that the convexified map is the true indifference map for each individual, then each individual winds up on the convexified curve I_0. Since this curve is flat at the point A, the price ratio is determined by the slope of BC. Now return to the individual, who has the original indifference curve I_0. At these prices, he will maximize utility at two different points, B and C, but not at any point in between. If, for example, A is half-way between B and C, then market equilibrium is realized by having half the individuals at B and half at C. If A is two-thirds of the way from B to C, the market equilibrium is realized by having two-thirds of the individuals buy the bundle C and one-third the bundle B. Note that each individual is at a point of maximum utility for him subject to his budget constraint, so that this is truly a competitive equilibrium and therefore efficient. (The earlier reference to "approximate equilibrium" is relevant when there are not enough individuals to split them in the right proportions between B and C. Thus, if A is .71 of the way from B to C and there are only 50 individuals in the economy, there should be 35½ individuals at C and 14½ at B. Thus, at C or B the discrepancy between supply and demand cannot be reduced below half an individual. This is a relatively minor discrepancy between supply and demand.)

Thus nonconvexity implies the existence of distinct *niches* for economic agents in a sense of the word which I take to be close to that used in ecology. One observes agents, identical in their economic data, engaged in diverse consumption patterns or other economic activities. Any given agent

may be indifferent between several of these niches, but equilibrium requires their coexistence. This argument underlies Adam Smith's discussion of specialization as opposed to Ricardo's, which was based on differences in the productivities of individuals or nations; it has been made explicit in an important but neglected paper of Houthakker.[7]

Let me now apply these abstract concepts to racial discrimination. We take up a model due to Becker and in a different form to Welch and not analyzed above.[8] Now we locate the discriminatory tastes in the W workers who are perfect substitutes for the B workers. To keep matters as simple as possible, assume there is only one kind of labor. Then, analogous to the assumption made about complementary forms of labor, we now assume that W workers have an indifference map between wages and the proportion W, so that at equilibrium, there is a relation,

$$w_W = w_W(W/L), \tag{11}$$

where w_W decreases as W/L increases from 0 to 1. As part of profit maximization, the firm will certainly seek that combination of W and B which will minimize the cost of hiring whatever total number of workers, $W + B = L$, it does hire. This cost is

$$C(W, B) = w_W(W/L)\, W + w_B B. \tag{12}$$

But it is easy to see that a firm will always achieve minimum cost with either an all-W or an all-B labor force. The two might be equally cheap, but certainly any combination with W and B both positive will be more costly than at least one extreme case and possibly more costly than both. To see this, consider two cases:

[7] See Hendrick S. Houthakker, "Economics and Biology: Specialization and Speciation," *Kyklos*, 9 (1956), 181-187.

[8] See for example, Gary Becker, *op. cit.*; and Finis Welch, "Labor Market Discrimination: An Interpretation of Income Differences in the Rural South," *Journal of Political Economy*, 75 (1967), 225-240.

(a) $w_W(1) \geqq w_B$: Recall that $W/L = 1$ means an all-W labor force. Then for any W, $0 < W < L$, $w_W(W/L) > w_W(1) \geqq w_B$, and therefore $w_W(W/L)W + w_BB > w_B(W+B) = w_BL$; hence, an all-$B$ labor force, with cost w_BL, is cheaper than the mixture. An all-W labor force has a cost $w_W(1)L$ and then $w_W(1)L \geqq w_BL$, so the all-W labor force is at any rate no less costly; if $w_W(1) = w_B$, the two extreme cases are equally cheap.

(b) $w_W(1) < w_B$: Then if $W < L$, $w_W(W/L)W + w_BB \geqq w_W(1)W + w_BB > w_W(1)(W+B) = w_W(1)\ L$; the all-$W$ labor force is cheaper than any other labor force.

Hence, if $w_W(1) > w_B$, every firm will find it cheapest to select an all-B labor force, and if $w_W(1) < w_B$, every firm will minimize cost by hiring an all-W labor force. But equilibrium requires full employment of both types of workers. The equilibrium then requires $w_W(1) = w_B$. But even then no firm will hire both W and B workers. At equilibrium every firm is segregated, but then the only observed wage for W workers is for those in all-W firms, i.e. $w_W(1)$, which is equal to w_B. Therefore, discrimination by W workers will not result in market wage differentials but instead does result in segregation.

In technical terms, the function, $C(W, B)$ is not a convex function, specifically, the isocost curves in W-B space are not concave to the origin. Convexity implies a tendency to the middle, to compromise; but here we have a rushing to extremes. We also have the characteristic implication of nonconvexity, a dispersion of firms with basically identical market opportunities into discrete niches.

Now, in going back over the analyses of Section 2, it can be observed that the case just discussed, of discriminatory tastes by a perfect substitute group of workers, is strikingly similar to that of discriminatory tastes by workers of a complementary type, the "foremen" of our example. Though a detailed analysis of the nonconvexities in this case has not yet been made, it is clear that the profit function defined by (9) is not in general a concave function. Rather, the surface

defined by profits as a function of W and B has holes scattered through it. Hence, it is at least possible that for certain values of w_W and w_B and some equilibrium relation, $w_F(W/L)$, there are several distinct points of maximum profits. Equilibrium on the three labor markets (W, B, and F) may be achieved by different amounts of these, even though each firm has the same production function and each faces the same wages for W and B and the same relation between w_F and W/L. Thus, there will be a partial segregation by firms.

The relation (10) stated earlier still holds for each firm, so the previous conclusions remain valid.

We can now reconsider the theory of employer discrimination. The utility function, $U(\pi, B, W)$, depends, it has been assumed, only on the ratio B/W. But it is shown in the Appendix that such a utility function cannot possibly have convex indifference surfaces everywhere.[9] Therefore it is possible and in fact likely that in the short run equilibrium will require the coexistence of firms of different sizes with different W/B ratios, even if all firms have the same utility function. Thus at least partial segregation is a likely outcome of the utility-maximization theory. All the firms will have to have the same utility, so that the larger firms will be those with the larger proportions of W workers since utility increases with π and with W/B.

In the long run, indeed, it can be seen that with constant returns to scale, there must be perfect segregation and equality of wages. For suppose there is not perfect segregation, i.e. there is at least one firm hiring both B and W workers. For that firm, $d_B > 0$ and $d_W < 0$. Since MP_L is constant in the long run, it follows from the preceding section that all B workers will be in firms with the smallest d_B and all W workers in firms with the (algebraically)

[9] The nonconvexity that arises when only ratios matter is of importance in the theory of pollution also. For the characteristic situation there is that the pollutee is faced with consuming air or water in which the proportion of pollutants is given to him.

smallest d_W. Then all firms have the same d_B and same d_W and therefore all have the same W/B ratio. The equilibrium values of w_W and w_B will be $MP_L - d_W$ and $MP_L - d_B$ respectively, where MP_L is the long-run marginal product of labor, a constant. But then any firm which increases its B/W ratio slightly can make positive profits; by increasing its scale, it can make indefinitely large profits with only a slightly altered W/B ratio. It would therefore have a higher utility, so we have a contradiction to the existence of equilibrium with at least one integrated firm. Hence, all firms are segregated. In an all-B firm, $d_B = 0$, for an increase in B does not change the W/B ratio and therefore leaves utility unchanged. Similarly, in an all-W firm, $d_W = 0$. It follows that, as in the case of discrimination by substitutes, the long-run equilibrium is one of perfect segregation and equal wages.

The corresponding analysis for discrimination by complementary employees has not been worked out. It can be conjectured, though, that segregation plus possibly competition among foremen with varying discriminatory tastes will greatly weaken wage differentials.

3. COSTS OF ADJUSTMENT

Utility-maximization theories, then, provide a coherent and by no means unreasonable account of the effect of discriminatory tastes on the market in the short run. Yet they become unsatisfactory in the long run. I propose as a possible explanation that long-run adjustment processes do not work as perfectly as they are usually assumed to. When there are significant nonconvexities, the adjustment processes called for must be very rapid indeed; marginal adjustments are punished, not rewarded. In the case of discrimination by substitute workers, the firm would have to be willing to fire its entire all-W labor force and replace it by an all-B labor force or vice versa in response to a very small change in wages. It is not unreasonable to assume that there

are costs, not ordinarily taken account of, which will restrain the firm from being quite so free in its adjustment behavior.

Now the idea that adjustment is costly has appeared in several diverse areas of economics. The costs of growth enter explicitly in some versions of the dynamic theory of the firm.[10] That is, the growth of the firm imposes a cost which depends on the rate of growth and which is additional to the purchase of capital goods.

The same principle, that capital costs of an unconventional kind play an important role in economic behavior and decisions, has been applied to the study of labor turnover, a problem more closely connected with ours. Operations researchers, in trying to draw up plans for hiring personnel, have incorporated in their models a fixed cost of hiring an individual. Sometimes it is also held that there is a cost attached to firing as well. These costs are partly in administration, partly in training. Even workers who have already been generally trained in the kind of work to be done must learn the ways of the particular firm. This approach, it has been argued by some, has important general economic implications; it implies that firms should not adjust their labor force to cyclical shifts in demand, since they then may incur both hiring and firing costs, costs that are avoided if the worker is retained during slack periods. Workers are being held in employment even though they contribute little to output to avoid the costs of rehiring them in the expected future boom. I do not know whether this explanation is in fact adequate but merely note that it is seriously considered.

A similar consideration may well explain why the adjustments which would wipe out racial wage differentials do not occur or at least are greatly retarded. We have only to as-

[10] See Edith Penrose, *The Theory of the Growth of the Firm* (Oxford: Oxford University Press, 1959); and Robin Marris, *The Economic Theory of 'Managerial' Capitalism* (Cambridge: Cambridge University Press, 1964).

sume that the employer makes an investment, let us call it a *personnel investment,* every time he hires a worker. He makes this investment with the expectation of making a competitive return on it; if he himself has no discriminatory feelings, the wage rate in full equilibrium will equal the marginal product of labor less the return on the personnel investment. Let us consider the simplest of the above models, that of discrimination by fellow employees who are perfect substitutes. If the firm starts with an all-*W* labor force, it will not find it profitable to fire that force, in which its personnel capital has already been sunk, and hire an all-*B* force, in which a new investment has to be made, simply because *B* wages are now slightly less than *W* wages. Of course, if the wage difference is large enough, it does pay to make the shift.

Obviously, in a situation like this, where there are costs to change, history matters a good deal. A fully dynamic analysis appears to be very difficult, but some insight can be obtained by study of a very special case. I here present only the results; the argument will be found in an earlier paper.[11] Suppose initially there are no *B* workers in the labor force. Then some enter; at the same time, there is an additional entry of *W* workers, and some new equilibrium emerges. Under the kinds of assumptions we have been making, a change, if it occurs at all, must be an extreme change, but there are now three kinds of extremes, or corner maxima. The typical firm may remain segregated *W*, though possibly adding more *W* workers; it may switch entirely to a segregated *B* state; or it may find it best to keep its present *W* workers while adding *B* workers. In the last case, of course, it will have to increase the wages of the *W* workers to compensate for their feelings of dislike; but it may still find it profitable to do so because replacing the existing W workers by *B* workers means wasting a personnel investment. If we stick closely to the model with all of its artificial conditions, we note that only the all-*W* firms are

[11] Arrow, "Some Models of Race in the Labor Market," Section E.

absorbing the additional supply of *W* workers, so that there must be some of those in the new equilibrium situation. On the other hand, there must be some firms that are all-*B* or else some integrated firms whose new workers are *B*'s in order to absorb the new *B* workers. It can be concluded in either case, however, that there will always remain a wage difference between *B* and *W* workers in this model. Further, there will be some segregated *W* firms. Whether the remaining firms will be segregated *B* or integrated depends on the degree of discriminatory feelings by *W* workers against mixing with *B* workers.

I have not worked out the corresponding analysis for the case where there are several types of workers with different degrees of discriminatory feelings against racial mixtures in the complementary types. Nevertheless, one surmises easily that similar conditions will prevail.

The generalization that may be hazarded on the basis of the discussion thus far can be stated as follows. If we start from a position where *B* workers enter an essentially all-*W* world, the discriminatory feelings by employers and by employees, both of the same and of complementary types, will lead to a difference in wages. The forces of competition and the tendency to profit-maximization operate to mitigate these differences. However, the basic fact of a personnel investment prevents these counteracting tendencies from working with full force. In the end, we remain with wage differences coupled with tendencies to segregation.[12]

4. IMPERFECT INFORMATION

There is an alternative interpretation of employer discrimination. It can be thought of as reflecting not tastes but perception of reality. That is, if employers have the preconceived idea that *B* workers have lower productivity than *W* workers, they may be expected to be willing to hire them

[12] The preceding five paragraphs have been quoted, with minor alterations, from Arrow, "Models of Job Discrimination," pp. 94-96.

only at lower wages. (Phelps has independently introduced a similar thesis.)[18] One must examine in detail the conditions under which this argument can be maintained, that is, the conditions under which the effects of these preconceptions are the same as those of discrimination in the strict sense of tastes.

First, the employer must be able to distinguish *W* workers from *B* workers. More precisely, the cost of making the distinction should be reasonably low. An employer might derive from his reading the opinion that an employee with an unresolved Oedipus complex will be disloyal to him as a father-substitute; but if the only way of determining the existence of an unresolved Oedipus complex is a psychoanalysis of several years at the usual rates, he may well decide that it is not worthwhile for him to use this as a basis for hiring. Skin color and sex are cheap sources of information. Therefore prejudices (in the literal sense of pre-judgments, judgments made in advance of the evidence) about such differentia can be easily implemented. School diplomas undoubtedly play an excessive role in employer decisions for much the same reason.

Second, the employer must incur some cost before he can determine the employee's true productivity. If the productivity could be determined costlessly, there would be no reason to use surrogate information, necessarily less valid even under the most favorable conditions. I suppose, therefore, that the employer must hire the employee first and then incur a personnel investment cost, as discussed in the last section, before he can determine the worker's productivity. This personnel investment might, for example, include a period of training, only after which is it possible to ascertain the worker's productivity; or indeed it may be only a period of observation long enough for reliable determination of productivity. In the absence of a personnel investment cost, after all, the employer could simply hire

[18] See Edwin S. Phelps, "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62 (1972), 659-661.

everyone who applied and fire those unqualified, or pay them according to productivity.

Third, it must be assumed that the employer has some idea or at any rate preconception of the distribution of productivity within each of the two categories of workers.

The simplest model to bring out the implication of these assumptions seems to be the following. Suppose there are two kinds of jobs, complementary to each other, say unskilled and skilled. All workers are qualified to perform unskilled jobs, and this is known to all employers. Only some workers, however, are *qualified* to hold skilled jobs. The employers need make no personnel investment in hiring unskilled workers but must make such an investment for skilled workers. The employer cannot know whether any given worker is qualified; however, he does believe that the probability that a random W worker is qualified is p_W and that a random B worker is qualified is p_B. An employer will eventually know whether or not a worker hired for a skilled position is in fact qualified, but this information is not available to other employers. He thus can count on keeping the qualified workers he hires.

Let r be the necessary return per worker on the personnel investment for skilled jobs. If a W worker is hired, then with probability p_W he is qualified; his productivity is MP_S, the marginal productivity of skilled workers, but the employer must pay a wage, w_W, so that the net gain to the employer is $MP_S - w_W$. On the other hand, if the worker hired turns out to be unqualified, the employer receives nothing. Hence, the expected return to a W worker hired is $(MP_S - w_W)\, p_W$. If the employer is risk-neutral, this must be equal to r. Similarly,

$$r = (MP_S - w_B)\, p_B, \tag{13}$$

and therefore,

$$w_W = q\, w_B + (1 - q)\, MP_S, \tag{14}$$

where $q = p_B/p_W$. Thus, if, for any reason, $p_B < p_W$, w_W is

a weighted average of w_B and MP_S and therefore lies between them; since from (13) we must have $w_B < MP_S$ (in order that the employer recoup his personnel investment), it follows that $w_W > w_B$, i.e. the effect of the differential judgment as to the probability of being qualified is reflected in a wage differential.

If there are price rigidities which prevent w_B from falling much below w_W, the same forces may be reflected in a refusal to hire B workers at all for skilled jobs.

Once we shift the explanation of discriminatory behavior from unanalyzable (or at any rate unanalyzed) tastes to beliefs, we are led to seek to explain these beliefs. One possible explanation runs in terms of theories of psychological equilibrium, of which Festinger's theory of cognitive dissonance is one of the most developed.[14] The argument is that beliefs and actions should come into some sort of equilibrium; in particular, if individuals act in a discriminatory manner, they will tend to acquire or develop beliefs which justify such actions. Hence, discriminatory behavior and beliefs in differential abilities will tend to come into equilibrium. Indeed, the very fact that there are strong ethical beliefs which are in conflict with discriminatory behavior will, according to this theory, make the employer even more willing to accept subjective probabilities which will supply an appropriate justification for his conduct.

Finally, one can also seek explanations in which p_W and p_B differ in reality, even though the intrinsic abilities of W and B workers are identical. Such an explanation requires some further assumptions. Specifically, whether or not a worker is qualified is now taken to be the result of a decision by him, rather than some type of intrinsic ability. More specifically, a worker becomes qualified by making some type of investment in himself. In accordance with the previous assumptions, this investment must not be observable

[14] For a more theoretical analysis see Leon Festinger, *A Theory of Cognitive Dissonance* (Evanston, Ill.: Row, Peterson, 1957).

by the employer. Hence, the investments are not the usual types of education or experience, which are observable, but more subtle types of personal deprivation and deferment of gratification which lead to the habits of action and thought that favor good performance in skilled jobs, steadiness, punctuality, responsiveness, and initiative.

Finally, it must be assumed, as is reasonable, that the human capital needed to qualify cannot be acquired on a perfect capital market. It follows that the proportion of either group (W or B) who qualify is an increasing function of the gain from qualifying. In accordance with our basic assumption that there is no intrinsic productivity difference between W and B workers, we assume that the supply schedules for the two groups are the same. Specifically, let $v_W = w_W - w_U$ be the gain to a W worker from qualifying, where w_U is the wage rate for unskilled labor, and similarly let $v_B = w_B - w_U$. Then we postulate an increasing function, $S(v)$, such that,

$$p_W = S(v_W), p_B = S(v_B). \tag{15}$$

Let MP_U be the marginal productivity of unskilled labor, so that,

$$MP_U = w_U. \tag{16}$$

Note that MP_S and MP_U are determined by the supplies of skilled and unskilled labor and these in turn are determined by the proportions p_W and p_B. Hence, the system consisting of the equations (13), (15), and (16) plus the equation obtained from (13) by replacing B with W constitute a system of equations in the unknowns w_W, w_B, w_U, p_W and p_U.

From the symmetric formulation of the system, it is clear that there can easily be a symmetric, nondiscriminatory equilibrium, i.e. one in which $p_W = p_B$ and $w_W = w_B$. Two questions can be raised: (1) can there be other, discriminatory, equilibria? (2) is the symmetric equilibrium stable? It can be shown that (1) in fact discriminatory equilibria

are bound to exist, and (2) the stability of the symmetric equilibrium depends on the parameters of the problem.

(1) The multiplicity of equilibria can be seen most easily if we assume that the black labor force is small compared with the white, so that any variations in black response have only small effects on the marginal productivities of skilled and unskilled labor and therefore only small effects on the wages of white skilled and of unskilled labor and therefore on the proportion of whites who are qualified.

From (15), (13), (14), and the definition of v_B, the basic equilibrium relation for the black labor force is,

$$(17) \qquad p_B = S(MP_S - MP_U - (r/p_B)) = S_B(p_B).$$

This equation is to be solved for p_B; we shall argue it is reasonable that there will in general be more than one solution with $p_B = 0$, as well as an equilibrium at $p_B = 0$. Since we are assuming that the effect of the black choice of p_B on MP_S and MP_U is negligible, the right-hand side of (17) depends on p_B alone. The two sides of (17) are graphed in Figure 2 below. The proportion qualifying can, of course, never exceed 1, and we may suppose that no wage difference within the range considered brings it up to 1; hence S_B, the proportion of blacks qualifying, is less than p_B for p_B close to 1. On the other hand, as p_B tends to zero, the wage differential between the skill levels,

$$v_B = MP_S - MP_U - (r/p_B),$$

tends to $-\infty$. There will surely be some wage differential which will cause a zero supply of qualified labor; even if skilled jobs are very attractive, so that there will be some supply even with a negative differential, the supply will surely disappear if the wage differential becomes a sufficiently large negative quantity, for example, if wages for skilled labor become close to zero. Hence, the S-curve in Figure 2 has roughly the shape indicated. Equation (17) is satisfied when the S-curve intersects the 45° line; hence, from the diagram it is clear that if there are any intersec-

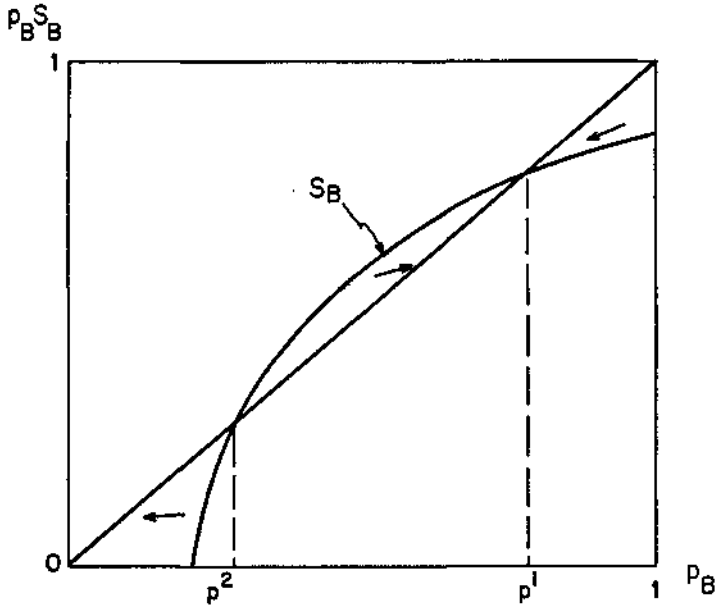

FIGURE 2

tions at all, there is more than one, so that multiple equilibria may be expected.

What happens when $p_B = 0$? Since $S_B = 0$ for p_B sufficiently small, as just argued, then $S_B = 0$ when $p_B = 0$. Hence, there must be at least three possible equilibria.

Notice that Figure 2 could also be interpreted as the diagram for analyzing the qualifying propensities of white labor force, since we are assuming that the supply function is the same for both. Hence, we see it is perfectly compatible with equilibrium conditions that p_W is at the highest intersection of the S-curve with the 45° line, while p_B is at a lower level or even zero. Thus, discrimination due to differing performance is possible even though the underlying assumptions are symmetrical with respect to race.

To discuss the plausibility of this situation, we must look into the stability of the alternative equilibria. First, we con-

fine our attention to the case already assumed, where the black labor force is small compared with the white, so that the marginal productivities of the two kinds of labor are independent of the behavior of the black labor force. To discuss stability, it is necessary to specify the dynamic model more precisely. We suppose that for given p_W and p_B, short-run equilibrium works itself out so quickly as to be instantaneously achieved. The basic dynamics then are Marshallian. That is, if the desired supply at any moment exceeds the current proportion qualified, the latter will increase and vice versa. In symbols,

$$dp_B/dt = k[S(v_B) - p_B]. \tag{18}$$

Then the movements of p_B are those indicated by the arrows in Figure 2; the highest possible equilibrium proportion qualified, p^1, is a stable equilibrium, and so is the value $p_B = 0$, but the intermediate value, p^2, is unstable.

If we venture a historical surmise, we may suppose that white workers started with a sufficiently large number of qualified individuals so that their proportion tended to p^1; but the black workers, starting from slavery with a low proportion qualified—a proportion correctly recognized by employers—drifted even lower toward zero or at least to some very low level. (It is unnecessary to caution the reader that this model is a gross simplification, intended to dramatize and make more extreme some existing tendencies, not to represent them literally.)

(2) We now argue that the nondiscriminatory equilibrium may be unstable. Here, we use the dynamic assumption (18) together with the same assumption for p_W; we no longer assume that the effects of black labor reactions on the whole labor market are negligible. Intuitively, we may consider a possible sequence of events, in which initially p_W slightly exceeds p_B for some reason. Then w_W slightly exceeds w_B and therefore, from (15), p_W tends to rise relative to p_B, therefore reinforcing the original disequilibrium. This verbal argument is certainly not conclusive nor very

convincing, and in fact the conclusion is valid only for some values of the parameters. We have to investigate the stability of the system defined by the pair of differential equations, (18), and the corresponding equation with p_B replaced by p_W, in the neighborhood of the nondiscriminatory equilibrium. While the algebra involved is elementary enough, there seems to be no way of making the result intuitively obvious. Hence, we simply reproduce the stability condition here, referring the reader for proof to an earlier paper.[15]

Let w_S be the common value of w_W and w_B at the nondiscriminatory equilibrium, p the common value of p_W and p_B. Then $v = w_S - w_U$ is the common value of v_W and v_B. Let E be the elasticity of $S(v)$ with respect to v, computed at the symmetric equilibrium value of v. From (13), $MP_S - w_S = r/p$ at the symmetric equilibrium; it is the excess of marginal product over wages for skilled workers. Then the condition for stability turns out to be that,

$$E(MP_S - w_S)/(w_S - w_U) < 1.$$

As might be expected, the greater the elasticity of the supply schedule for qualified labor, the more likely is the system to be unstable. Similarly, the greater the difference between marginal product and wage for skilled workers, the more likely is instability; this difference would be zero if there were no personnel investment costs for skilled workers, and then the system would certainly be stable. Finally, and less intuitively, the larger the wage gap between the two types of labor, the less likely is instability.

I believe these results are only the barest fragment of what could be found with better and more detailed systems in which there is an interaction between reality and perceptions of it. One must consider still more precisely how individual employers acquire knowledge which will modify their initial estimates of distributions as differing between

[15] Arrow, "Some Models of Race in the Labor Market," Section F.

groups and in turn the effects of these perceptions on the market and therefore on any incentives to modify those abilities.

APPENDIX

Nonconvexity of Indifference Maps Depending on Ratios

We suppose that employer discrimination is determined by a utility function, $U(\pi, B, W)$, where multiplying both B and W by the same positive constant leaves utility unchanged. We also assume that U is an increasing function of profits, π. It will be shown that the indifference map defined by U cannot have convex indifference surfaces; specifically, a convex combination of two indifferent points is not everywhere at least as good as either.

Choose any point (π_0, B_0, W_0). Then choose $\pi_1 < \pi_0$ (as close as needed) and B_1, W_1 so that

$$U(\pi_1, B_1, W_1) = U(\pi_0, B_0, W_0). \tag{1}$$

From the assumptions made, (1) will continue to hold if B_1 and W_1 are reduced in the same proportion. Hence, B_1 and W_1 can be chosen arbitrarily small.

If the indifference map defined by U has everywhere convex indifference surfaces, then the average of the two points must be at least as good as (π_0, B_0, W_0). That is,

$$U(\pi', B', W') \geqq U(\pi_0, B_0, W_0),$$

where $\pi' = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_0$, $B' = \frac{1}{2}B_1 + \frac{1}{2}B_0$, $W' = \frac{1}{2}W_1 + \frac{1}{2}W_0$. But then, since,

$$U(\pi', 2B', 2W') = U(\pi', B', W'),$$

we have,

$$U(\pi', 2B', 2W') \geqq U(\pi_0, B_0, W_0),$$

or, by definition,

$$U(\pi', B_1 + B_0, W_1 + W_0) \geqq U(\pi_0, B_0, W_0).$$

But B_1 and W_1 can be chosen as small as desired. Let them approach 0; by continuity,

$$U(\pi', B_0, W_0) \geqq U(\pi_0, B_0, W_0),$$

which is a contradiction to the assumption that U is increasing in π, since $\pi' < \pi_0$.