calculate the present value of the full set of benefits, simply divide each year's benefits (the higher earnings) by $(1 + r)^T$, If the benefits are denoted as $B_1, \ldots B_T$, then the present value is

$$PV(B_1, \ldots, B_T) = \tag{1}$$
$$B_1/(1 + r) + B_2/(1 + r)^2 + \cdots + B_T/(1 + r)^T = \sum_{t=1}^{T} \frac{B_t}{(1 + r)^t}$$

where Σ, the Greek letter sigma, is the summation sign. Note that benefits received in the first year are divided by $(1 + r)$, benefits in the second year are divided by $(1 + r)^2$, and so on, up to year T. The term at the end involving Σ is just a compact way to express a long sum of terms; the summation starts with $t = 1$ and goes through $t = T$. The summation thereby reproduces the longer addition list on the other side of the equals sign. This is a very important expression that you should know.

To evaluate how good an investment is, economists compute its **internal rate of return.** The internal rate of return is a measure of the net value or profitability of an investment that reflects how large the benefits are relative to the costs, after properly accounting for the different timing of the costs and benefits using present value analysis. To be precise, the internal rate of return to an investment is the interest rate at which the present value of all the benefits just equals the cost of the investment. It is the interest rate r^* that solves the following expression:[8]

$$\sum_{t=1}^{T} \frac{B_t}{(1 + r^*)^t} = C \tag{2}$$

The left-hand side of this expression is the present value of the benefits, using interest rate r^*. The right-hand side of the expression is the costs, which are assumed to be entirely incurred in the current time period; that's why we don't need to compute a present value for them. If the benefits are very large relative to the costs, then the internal rate of return will be large as well: It will take a high interest rate to reduce the present value of benefits to the smaller level of the costs. If the benefits are not that much larger than the costs, then

[8]In most equations, the variable that is being solved for (here, r^*) is all by itself on the left-hand side of the equation. Because this computation involves a long summation, that expression is impossibly complicated. Equation (2) shows the solution indirectly. As a practical matter, internal rates of return are almost always solved with a financial spreadsheet program that finds the solution by trial and error in an iterative procedure. To compute an internal rate of return with a spreadsheet program, set up a column of net benefits, treating the cost as a negative benefit. You will need to provide a starting value for the internal rate of return; the program then solves for the solution by trial and error. If the present value of benefits at the starting value is greater than the costs, the program recomputes present value using a slightly higher interest rate; if the present value of benefits at the starting value is negative, a lower interest rate is used. The process continues until an internal rate was found at which the present value of benefits equals zero.

the internal rate of return will be very small. An investment with a low internal rate of return is unlikely to be an attractive investment from an economic standpoint.

To determine whether any particular investment in human capital is worth making, an individual should compare the internal rate of return to the market interest rate. The market interest rate represents the rate at which an individual can borrow the funds to finance the investment or, alternatively, the rate at which he or she could invest the proceeds. The decision rule is simple: If the internal rate of return is greater than the market interest rate, then the investment makes economic sense. If the internal rate of return is smaller, however, there is no economic basis for proceeding. This means that anything that lowers the internal rate of return to an investment in human capital may cause an individual to forego the investment entirely. Could that apply to women and thereby cause them to invest in less human capital than men?

Gender Differences in Human Capital

THEORY—WHY THERE MIGHT BE GENDER DIFFERENCES IN HUMAN CAPITAL

Suppose a woman expects to take some time off from working in the labor market for family responsibilities. This was, as we have seen, certainly the norm in previous generations, and even today, it is certainly not uncommon. (See Box 3.1 on the division of labor within families for more information.) In those years, the benefits to any investment in human capital will be zero. Alternatively, if a woman works part-time to accommodate family responsibilities, her benefits will be positive, but smaller than they might otherwise be. These patterns of labor force participation lower the internal rate of return for two related reasons: First, total benefits are lower, and, second, the benefits that are lost are those with a high present value because they typically come relatively early in a woman's adult life.

Investment in Education Let's consider a specific example of how periods of not working in the labor market affect the internal rate of return. Figure 9.1 shows average earnings in 2006 by age for women between the ages of 18 and 64 with college and high school degrees who were year-round full-time workers. The data come from the Census Bureau's annual survey of individual earnings. The earnings are averages by five-year age brackets, which is why the two lines have flat regions and then sudden jumps. For high-school graduates, earnings start at age 18, while they start at age 22 for college graduates.

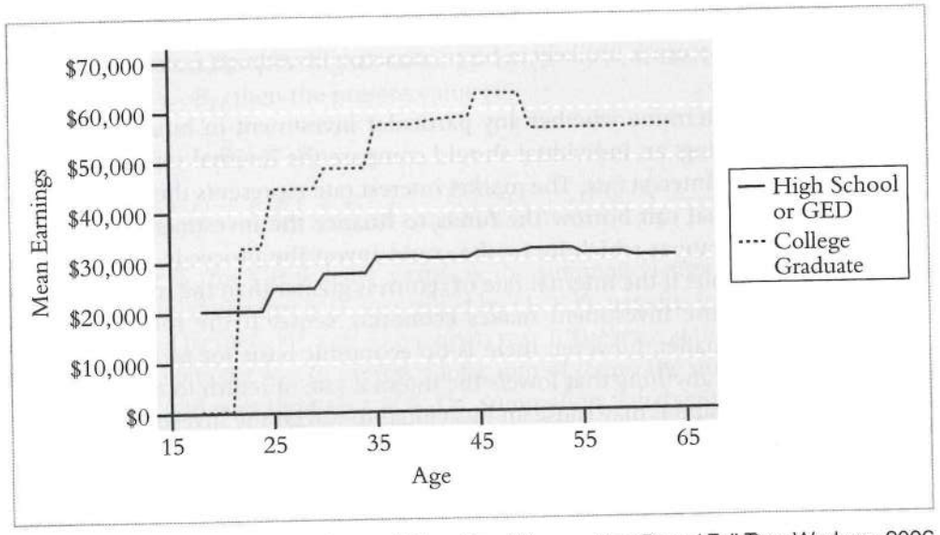


FIGURE 9.1 Mean Earnings by Age and Education, Women, Year-Round Full-Time Workers, 2006

We can use this data to compute the internal rate of return to a college degree for women under different assumptions about work. In doing that, we follow standard practice and assume that the foregone earnings during college are measured by what is *actually* earned by the high school graduates during those years (i.e., between the ages of 18 and 22). These are the opportunity costs. That is, admittedly, far from perfect, but it's probably in the ballpark. To that, we will need to add an estimate of direct costs. Similarly, the benefits can be measured by the difference in average earnings between women college graduates and women high school graduates. As you can see, the benefits are substantial. The earnings gap is about $20,000 at ages 25 to 34, $25,000 at ages 35 to 44, and peaks at more than $31,000 at ages 45 to 49.[9] Thus, Figure 9.1 shows the costs and benefits associated with full-time year-round work throughout a woman's career. The area between the two lines from ages 22 to 65 represents the benefits of a college degree, while the area between them at ages 18 to 22 are the opportunity costs of the foregone earnings.

The internal rate of return to a college degree is the interest rate at which the present value of discounted benefits just equals the opportunity costs plus any direct costs. Suppose out-of-pocket expenses for college are $15,000 per year for four years. Then, the internal rate of return to the earnings profile

[9]We also are assuming that current college and high school graduates will, when they are older, earn what is actually earned in 2006 by women at older ages. This assumption is not perfect, but it is good enough for our purposes.

shown in Figure 9.1 is a very respectable 11.5%. Using comparable data for men, the internal rate of return for college is 13.7%.

With rates of return like these, a college investment is a very good investment, indeed. What happens, though, if a woman takes time out of her career? If both the college graduate and the high school graduate left the labor market at age 25 and returned to year-round full-time work at age 29, the internal rate of return to the college degree falls 2.5 percentage points to 9.0%. If a woman leaves the labor market for ten years (from age 25 to age 34), the internal rate of return falls about another 1.6 points, all the way down to 7.4%. And with fifteen years out of the labor force, the internal rate of return is just 6.0%, just about half the value with full-time continuous work. Imagine how low the rate of return would have been for women born in the early decades of the twentieth century for whom market work often ceased with marriage. See Box 9.1 for evidence on the return to a college degree for women during the baby boom.

Box 9.1 The Return to a College Degree During the Baby Boom

Claudia Goldin provides estimates of the internal rate of return for women to a college degree in the mid-1950s.[10] The median female college graduate in that time period married in the year of her graduation, worked for four years, and then exited the labor force for about eight years, returning at age 35. For high school graduates, Goldin assumed that they, too, worked four years after school, took ten years off, and then returned to work. College-educated women earned about 30% more than high-school educated women at the beginning of their careers and about 40% later on. With further allowance for out-of-pocket college costs, Goldin estimated an internal rate of return for college graduation of about 5%, certainly no goldmine and about half of the internal rate of return for men at that time. She notes that the income premium to female college graduates was about the same as for men. Their lower internal rate of return was entirely a function of their briefer employment.

College was, however, somewhat more financially lucrative for women if viewed more broadly. Goldin notes that college attendance and especially college graduation greatly increased the probability that a woman married a college graduate. For example, only 10% of high school graduate women were married

[10]Claudia Goldin, "Career and Family: College Women Look to the Past" in *Gender and Family Issues in the Workplace*, edited by Francine Blau and Ronald Ehrenberg (New York: Russell Sage Foundation, 1997), pp. 20–58.

to a college graduate, compared to two-thirds of all college graduate women. And, of course, these more-educated men earned more than their less-educated counterparts. Goldin reports that, on average, the husband of a college graduate woman earned almost 40% more than the husband of a high school graduate woman. Taking all this into account, Goldin estimates that the "full" internal rate of return to college for women was in the 10 to 11% range rather than the 5% associated with its direct labor market impact.

The impacts on the internal rate of return are so large because the years of absence from the labor market are early career years with high present value. It is easy to see that leaving the labor market can make a huge difference in the internal rate of return of the investment. If the rate of return is low enough, it may not make economic sense to make the investment in the first place. And that, in turn, would result in women having less human capital than men.

Investment in on-the-Job Training As we discussed earlier, most jobs provide a worker with not only a wage rate, but also the opportunity to acquire human capital (learn skills) that will be valuable subsequently in terms of higher wages. This learning on the job is precisely what economists mean when they talk about on-the-job training. Jobs that provide OJT will typically pay lower current wages than jobs that offer no training, but compensate for that with higher wages after the training is received. In that way, acquiring OJT is a human capital investment with costs and benefits that can be measured by its rate of return.

In the case of general training, the skills learned are widely valuable. In that case, human capital theory shows that individuals receive *all* the benefits of training and pay *all* the costs. In the case of specific training, the skills learned have a more limited market, and in the extreme case of completely specific training, the skills are valuable only in the current firm. Completely specific training often reflects knowledge of procedures and/or people that are unique to a firm. University department coordinators are a good example of workers who have a great deal of knowledge of forms, people, and procedures that are highly university-specific. In the case of specific training, it turns out that firms and workers share the costs and the benefits of investment in training.

Precisely as in the case of education, it is possible to think about the rate of return to the investment in OJT. Suppose a job has a starting salary that is $2,500 lower than another otherwise similar job but provides more on-the-job training—enough more that your earnings would increase by $500 per year thereafter for as long as you stay on the job, starting in the next year. (Again,

this is a simplified arithmetic example; in the real world, the benefits might well increase over time.) If you use that on-the-job training for ten full years, the internal rate of return to the $2,500 investment is over 15%. If, instead, you use the training for only six years and then drop out of the labor market, the return drops below 5.5%.

In the case of specific training, a firm will want to make sure that any worker that they train will stay with the firm long enough that it can recoup its costs and earn a reasonable rate of return on its investment. So here, too, women's history of intermittent labor force participation may make firms more reluctant to offer them jobs with substantial specific training.

Differences in labor force participation are not the only thing that could lead to differences in internal rates of return. Suppose that, because of labor market discrimination, women don't earn as much as men for the human capital they possess. Suppose opportunities for advancement are limited, no matter how much human capital they have. In the OJT example, suppose a woman received only $400 more per year after training, rather than $500, because she wasn't promoted. Even if she expected to have the same ten-year career as a man, her rate of return would be just over 9.5%, not the 15% that a man might receive. Again, it is easy to see that if labor market discrimination lowered rates of return sufficiently, a woman might rationally choose not to make a costly investment in human capital. "What's the point?" she might reason. I can do nearly as well without making the investment.[11]

Let's summarize the main points about the human capital approach to gender earnings differences. An investment in human capital has a financial return that can be measured by the internal rate of return. Rates of return are affected by many factors, but especially by participation in the labor market and by discrimination, both of which affect the benefits received. Individuals make decisions about investment in human capital on the basis of the rate of return it offers. If women anticipate spending some years out of the labor force for family responsibilities, this will lower the internal rate of return to human capital investment. Under these circumstances, they might well rationally choose to invest in less human capital. The ultimate source of these potential labor market human capital differences by gender is the traditional organization of households, with its characteristic division of labor on the basis of gender, that we discussed in Chapter 3. Labor market earnings

[11]This kind of argument applies with great force in the case of race discrimination in the United States, South Africa, and elsewhere. Think about the human capital investment situation facing Black workers in the time period when overt discrimination was rampant and not illegal. They had very few incentives to acquire skills because they would be unlikely to reap the benefits. Note that this means that discrimination may cast a long shadow—its labor market effects are felt a long time after it may have been eliminated. An argument like this is sometimes used to support affirmative action programs.

depend on the amount of human capital a worker has. If women invest in less human capital than men, they will earn less than men.

EVIDENCE—GENDER DIFFERENCES IN HUMAN CAPITAL

The actual amount of human capital that an individual has is not directly observed. For example, it is not possible to say that a certain person has 17 units of human capital or 42 units or any other number. That doesn't mean that the concept isn't useful. Human capital acquired through schooling is usually assumed to be related to the number of years of completed schooling, while human capital in the form of on-the-job training is usually assumed to be associated with the number of years of work experience.

Years of education and years of work experience are by far the two most important factors used by economists to explain differences in wages among individuals and between men and women. If human capital differences are responsible for at least some portion of the gender gap in earnings, there must be differences between men and women in these measures. Let's look at the facts.

Gender Differences in Education In Chapter 8, we reviewed gender issues in educational attainment. We concluded there that women and men have graduated from high school in equal numbers since the 1940s at least, but that two very different gender gaps in college completion have existed at different times. Through at least the early 1990s, men were more likely to be college graduates than women. From the early 1960s through the late 1970s, this difference was large—about 7 to 8 percentage points (14.5% versus 7.0% in 1960, 20% versus 12% in 1970, and 28% versus 20% in 1979). In percentage terms, these differences are quite large, so it is clear that across the population in those years, men had more education than women. But thereafter, the gender gap in college completion first narrowed, then disappeared, and finally emerged in substantial favor of women. Since 1982, women have outnumbered men among graduating college students and since 1996 they have outnumbered men as college graduates among 25- to 34-year-olds; in 2007, the difference was more than 6 percentage points (34.2% versus 27.9%). By the mid-2000s, men's educational advantage has been completely eroded for the total working-age population. For younger age groups, women now thoroughly dominate in numbers in college and have at least rough parity or more with men in most professional degree programs. They still lag in the hard sciences at the postgraduate level.

These trends mean that educational differences between men and women could have played an important part in earnings differences through the 1970s and a declining but still genuine part through the early 1990s. Moving forward and especially among younger workers, it is unlikely that educational differences can be an important factor in explaining the gender gap in pay.

Gender Differences in Work Experience Historically, as we discussed in Chapter 7, women have been far less likely to be in the paid workforce. In the first third of the century, participation in the paid labor market was largely confined to women before marriage and to women who did not marry. The labor force participation rate for married women was just 10% in the early part of the century and only 20% in 1950. The gender gap in participation rates didn't begin to close until the 1960s and 1970s, and even now there is a 15 percentage point gender difference in participation rates. Thus, because women's labor force participation rates were substantially lower than men's, it follows as a matter of arithmetic, that women must on average have accumulated fewer years of work experience than men. It wouldn't be surprising if this contributed to the gender gap in earnings.

In fact, the gender difference in work experience is more complex than these time trends suggest. Claudia Goldin's research on labor force participation by married women, which we discussed in Chapter 7, showed an important changing pattern involving life cycle patterns of work. Table 9.1 is an extract from Table 7.3. It shows by age bracket the labor force participation rate of White married women for three cohorts of women, each born twenty years apart from 1886 to 1926. Given the ages and birth years involved, this table covers the labor market experience of married White women from 1920 (born 1886–1895, aged 25 to 34) to 1980 (born 1926–1935, aged 45 to 54). Look carefully at the entries in the first row for women born near the end of the nineteenth century. Not only was their labor force participation rate incredibly low when they were young married women (7.7%), but it increased very little as they aged. One way to think of these numbers—imperfect, but useful—is that 7.7% of the women worked essentially steadily from age 25 to 54; 2.1% (9.8% − 7.7%) entered the labor market between the ages of 35 and 44; and another 0.3%, between the ages of 45 and 54. Thought of that way, more than three quarters (7.7%/10.1%) of the women working between the ages of 45 and 54 in 1940 had been working continuously for the previous twenty years.

Year of Birth	Age 15–24	Age 25–34	Age 35–44	Age 45–54
1886–1895	n.a.	7.7	9.8	10.1
1906–1915	13.3	16.7	25.3	38.6
1926–1935	24.9	26.7	44.4	53.4

TABLE 9.1 Life Cycle Labor Force Participation Rates for White Women by Age and Birth Cohort

Source: Adapted from Goldin, Claudia. *Understanding the Gender Gap.* © 1990 by Oxford University Press, Inc. Reprinted by permission of the publisher. http://www.oup.com

The figures for the 1906–1915 birth cohort indicate that not only are participation rates higher, but they increase with age. This increase represents the phenomenon of labor market reentry by married women as they and, more importantly, their children aged. Thus, the participation rate increased almost 9 percentage points between those aged 25 to 34 and those aged 35 to 44 and increased another 13.3 percentage points over the next ten years. Applying the same reasoning to the 1906–1915 birth cohort that we used for the older cohort, we see that only 43% (16.7%/38.6%) of the women in the labor market at ages 45 to 54 in 1960 had been there twenty years before. This means that the average total work experience of this cohort of working women at ages 45 to 54 must have been lower than for the earlier cohort because so many more of them were reentrants who had relatively little work experience. For the cohort shown in the last row of the table, exactly 50% (26.7%/53.4%) of the women in the labor market at ages 45 to 54 in 1980 had been in the labor market twenty years previously. The average total work experience of this cohort of working women at ages 45 to 54 is therefore higher than in 1960, but still substantially lower than in 1940.

Goldin emphasizes the importance of this life cycle pattern of labor force participation for understanding the trend in the gender gap in wages. In the early part of the century, relatively few women worked, but those who did, worked more or less continuously, year after year. Thus, differences in years of work experience between *working* men and *working* women were relatively small. Later in the century, more women worked, but relatively few worked continuously. As a result, a greater proportion of women working at any point in time had previously been intermittent labor force participants. Thus, differences in years of work experience between working men and women were larger than in the past. Paradoxically, the increase in women's labor force participation, which took the form of reentry into the labor market, had the initial effect of increasing the gender gap in years of work experience among working women.

Goldin notes that this pattern meant that relatively few White married women established careers in the same way that men did. Rather, she writes "married working women who entered the labor force in their middle years had, by and large, entered a rather new world. ... They were not bringing to the labor market recently acquired skills. Many were, instead, reentering occupations they had left many years, or possibly decades, before, and most were inadequately prepared for their new labor market roles" (p. 23).

James Smith and Michael Ward have made a similar point in their research on the historical patterns of labor force participation by women.[12]

[12]James P. Smith and Michael Ward, "Women in the Labor Market and in the Family," *Journal of Economic Perspectives*, 3 (1) (1989): 9–23.

They computed the average accumulated years of labor market experience from 1920 to 1986 for *all* women and for the subset of women who were *working* in those years. Table 9.2 shows the figures for working women at ages 25, 35, and 45. In 1920, average work experience among working women at age 35 was almost 12 years. One way to appreciate this number is to realize that if work began with a woman's graduation from high school at age 18, the maximum work experience at age 35 would be seventeen years. So the average working woman at age 35 had worked more than two-thirds of potential years beginning at age 18. At age 45, average work experience was more than sixteen years, or almost 60% of the possible work years since age 18.

By looking down the columns, we can see how average work experience changed at each age between 1920 and 1986. Remember that this is a time period in which the labor force participation rate increased by more than 30 percentage points, from 24% to 55%. Average labor force work experience among working women, however, increased very little at all. At age 35, average work experience actually fell by 0.1 years between 1920 and 1980, before rising by about two years between 1980 and 1986. Average work experience at age 45 also rose very little, not quite three years in more than sixty years. These figures are consistent with the information that Goldin presented on life cycle labor force participation.

When Smith and Ward computed average work experience across *all* women, regardless of whether they were currently working, work experience increased much more across this time period because of the large increase in labor force participation. At age 35, women's average work experience increased more than six years and at age 45, it increased nearly seven years. In 1920, the difference between the average work experience of *working* women and *all* women was very large because many women almost never worked in

Year	Age		
	25	35	45
1920	5.6	11.8	16.4
1950	5.9	10.6	17.0
1980	6.2	11.7	18.3
1986	6.5	13.5	19.3

TABLE 9.2 Average Years of Work Experience for Working Women by Age and Year

Source: Adapted from James P. Smith and Michael Ward, "Women in the Labor Market and in the Family," Table 2. *Journal of Economic Perspectives*, 3(1) (1990). Copyright © 1990 by the American Economic Association. Reprinted by permission of the authors and the publisher.

the paid labor market. In that time period, the average work experience of *working* women was relatively high, but the average work experience of *all* women was quite low. By 1986, however, this had changed: The averages for the two groups were much more similar because a much higher proportion of women now worked in the labor market at some time in their lives.

From our standpoint, the most important point of all this research is that women have indeed, on average, worked fewer years in the labor market than men of the same age and that, paradoxically, the gap did not narrow when the labor force participation rate of women first began to increase in the 1950s through the 1980s. The rising labor force participation rate was largely the product of reentry into the labor market by women who had been absent for many years. Their presence thus prevented the gender gap in work experience from narrowing further.

In the next chapter, we will develop a technique to determine exactly how important education and work experience differences between men and women are in explaining the gender gap in earnings.

Labor Market Discrimination and Women's Earnings

An alternative explanation for the gender gap in earnings is that women earn less because they are discriminated against in the labor market. The standard economic analysis of discrimination is based on Gary Becker's *Economics of Discrimination*, a book written in the mid-1950s. Becker's analysis was developed in the context of racial discrimination, which was by far the most widely noted form of discrimination at that time; public awareness of possible labor market discrimination against women was still at least a decade or two in the future.

The application of Becker's approach to gender discrimination is complicated in at least one way. Typical discrimination models are based on aversion or dislike. Whatever we may individually feel about these dislikes and their legitimacy, we understand that individuals do sometimes harbor genuine negative feelings for persons of differing race, ethnicity, religion, sexual preference, and so on. But what about gender? It cannot really be said that men dislike women; after all, in the non-market sphere of life, they often like them very much. So here it is primarily a matter of context. If discrimination models are to make any sense, it must be that some men feel aversion to women holding particular employment positions, rather than having an across-the-board dislike. Bear that distinction in mind as we consider the models.

The economic approach to discrimination is often puzzling to students. First, it makes absolutely no effort to explain *why* discriminatory preferences exist. In that respect, it follows the tradition in economics of treating preferences as exogenous, that is, as a given of the situation being analyzed. For example, economists make no effort to understand why certain people like broccoli, while others prefer string beans or corn, although those preferences are certainly an important part of explaining the demand for and market prices of broccoli, string beans, and corn. The economic approach simply takes those likes and dislikes as givens, and then analyzes how, holding those preferences constant, demand is affected by incomes and prices.

Without a doubt, how people develop the preferences they have, particularly their discriminatory preferences, is an important issue. But it is not one that traditional economic analysis can shed much light on. Psychologists, anthropologists, and sociologists all have comparative advantages in analyzing that issue. The comparative advantage of economics is in analyzing how these preferences affect market outcomes.

Thus, in his work, Becker assumed that some individuals had what he called a **taste for discrimination.** This may seem like an odd choice of words, but it is part of Becker's effort to place likes and dislikes on the basis of gender, race, and other personal characteristics within the general economic framework of preferences. A taste for discrimination is not just a like or dislike but rather a like or dislike that an individual is willing to act on. In Becker's words: "If an individual has a taste for discrimination, he [sic] must act as if he were willing to pay something, either directly or in the form of a reduced income, to be associated with some persons rather than others" (p. 14).

Second, Becker's analysis of discrimination is embedded within the conventional economic analysis of utility-maximizing individuals and profit-maximizing firms, operating in markets that often constrain their behavior. Becker's approach therefore focuses on market equilibriums, and especially long-run competitive equilibrium.

Becker examined three potential sources of discrimination—employers who had tastes for discrimination regarding their employees (**employer discrimination**), employees who had tastes for discrimination regarding their co-workers (**employee discrimination**), and customers who had tastes for discrimination regarding the suppliers with whom they interacted (**customer discrimination**). We examine each in turn.

EMPLOYER DISCRIMINATION

Let there be two kinds of workers, whom we call m and f for male and female. To isolate pay differences resulting from discrimination, assume that the two groups of workers are identical in all relevant respects. Technically, they are

perfect substitutes in production, exactly and identically productive, so no employer would have any reason to favor one group over the other in the absence of discrimination. Indeed, in the absence of discrimination, the wages for the two groups of workers would be identical because otherwise no employer would ever choose to hire the more highly paid group.

But what happens if employers have a taste for discrimination for (are prejudiced against) f workers. (It probably makes sense to think of the employers as members of group m). To give this prejudice a quantitative measure, Becker introduced what he called a **discrimination coefficient.** It operates in the following specific way. Let d stand for the discrimination coefficient, w for the money wage that an employee receives, and W for the gross (or total) wage as perceived by the employer. For an employer with discrimination coefficient d, $W = w \times (1 + d)$. The gross wage now consists of two parts—the money wage (w) and the monetary equivalent of the taste for discrimination ($w \times d$). For example, if the money wage were \$5 and $d = 0.25$, the gross wage would be $\$5 \times 1.25 = \$6.25$. Facing a money wage of \$5, this employer would act as if the wage were \$6.25, composed of \$5.00 in money wages and a disutility cost equal to \$1.25. A more prejudiced employer would have a higher value of d and thus a higher gross wage for any given money wage.[13] An employer with no prejudice at all would have $d = 0$.

How do discriminatory preferences affect the equilibrium wages of men and women? We begin with a simple case in which all employers have the same taste for discrimination, which we denote by d'. Let w_m stand for the money wage rate for men and w_f be the money wage rate for women. To get the gross wage for women, as perceived by the employer with a taste for discrimination, we multiply that wage by $(1 + d')$, getting $w_f \times (1 + d')$. If $w_m < w_f \times (1 + d')$, then women are more expensive (inclusive of all costs) than men, so that all employers would choose to hire only men. If $w_m > w_f \times (1 + d')$, just the opposite is true—the men are too expensive and even the prejudiced employers would want to hire only women because the lower wage more than compensates for the higher psychic costs of hiring the women. These employers are prejudiced, but they are also capitalists!

The only way both groups can find employment is when their gross wages are equal. Thus, in equilibrium, the relative wages for men and women must satisfy this relationship:

$$w_m^* = w_f^* \times (1 + d') \tag{3}$$

[13]Note that the discrimination coefficient is multiplicative, not additive. If it were additive, so that the net wage was $w + d$, the same general results would hold.

where the asterisk (*) means this is an equilibrium wage rate. We can rewrite this as a wage ratio:

$$w_f^*/w_m^* = 1/(1 + d') \tag{4}$$

Equation (4) shows that the women's equilibrium wages will be less than the men's, as long as $d' > 0$. If, for example, $d' = 0.50$, then $w_f^*/w_m^* = 1/1.5 = 67\%$. If $d' = 0.25$, then $w_f^*/w_m^* = 1/1.25 = 80\%$. Equations (3) and (4) don't establish what the actual wages for men and women will be, just the relationship between the wages. The absolute wage level will depend on general worker productivity.

Now let's make the problem a bit more realistic and also a bit more complex. Instead of assuming that all employers have the same discrimination coefficient, suppose the discrimination coefficient ranges from 0 (no taste for discrimination) for some employers to some large number that we'll call d_m; the m subscript stands for the maximum. We use d_j to stand for the discrimination coefficient for each particular employer. Just as before, each employer will compare the gross wages of men and women and hire the less expensive group. When $w_m = w_f$, employers with $d = 0$ will be the only ones willing to employ women, because for everyone else, the gross wage of women $[w_f \times (1 + d_j)]$ exceeds the gross wage for men (w_m). If w_f were just a bit lower, firms with the next lowest value of d_j would be just willing to employ women, and so on. As w_f continues to fall, more and more employers are willing to employ women, until the wage is so low that even the employer with $d = d_m$ would be willing to hire women.

Table 9.3 shows how this might work, for employers A, B, C, D, and E, arranged from lowest to highest discrimination coefficient. If $w_f = \$10.00$, the employers will view that wage as if it were \$10.00 (employer A), \$12.50 (employer B), and so on up to E, who acts as if the wage were \$100! In

Employer	d	Net Wage $[w \times (1 + d_j)]$, If w_f = \$10.00	Maximum Wage Willing To Pay Women, If w_m = \$10.00
A	0	\$10.00	\$10.00
B	0.25	\$12.50	\$8.00
C	1	\$20.00	\$5.00
D	3	\$40.00	\$2.50
E	9	\$100.00	\$1.00

TABLE 9.3 Employer Discrimination Coefficients and Willingness to Hire

each case, the net wage is just $w \times (1 + d_j)$. The final column shows the wage at which they would be just willing to hire women employees, assuming that the male wage was \$10.00. This is the value of w_f, such that $w_f \times (1 + d_j) = \$10.00$; rearranging, this wage is equal to $\$10.00/(1 + d_j)$. It ranges from a high of \$10.00 (employer *A* who doesn't care whom he/she hires) to \$1.00 for employer *E*, who clearly cares an enormous amount.

Figure 9.2 illustrates this situation. Suppose, just to simplify a bit, that each of the five firms were willing to hire a maximum of ten workers and that the wage for men was \$10.00. What would the demand curve look like? If $w_f = \$10.00$, firm *A* is willing to hire anywhere between 0 and 10 women. (At that wage, it is indifferent between hiring men and hiring women, and ten workers is the maximum). That corresponds to the top short horizontal section of the demand curve. If the wage were \$8.00, firm *A* would definitely hire ten women (it would actually do so if the wage were \$9.99) and now firm *B* is willing to hire between one and ten women. So the demand curve is a horizontal line stretching from 10 to 20 when $w_f = \$8.00$. At a wage of \$5.00, firms *A* and *B* will hire a total of twenty women, and firm *C* will hire anywhere from one to ten. And so on, down to firm *E*, whose demand for women

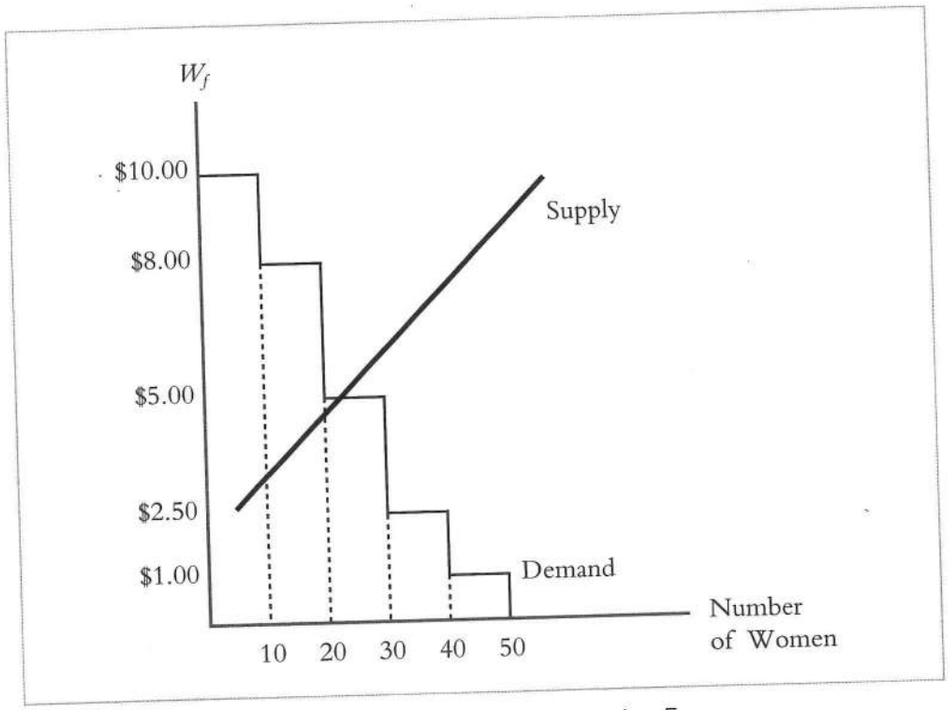


FIGURE 9.2 The Demand For Women Workers, Employers *A* to *E*

workers kicks in when the wage for women is \$1.00. The demand curve looks like a descending set of stairs.

The actual wage in the market will depend on where the supply curve intersects this strange demand curve. For the supply curve drawn in Figure 9.2, the equilibrium wage will be \$5.00. At this wage, firms *A* and *B* hire only women, firms *D* and *E* hire only men, and firm *C* hires both. If the supply were smaller, the wage would be higher, because then the women would need to deal with the less-prejudiced employers only.

If there are many firms instead of five, and if the value of d_j differs just slightly from one firm to the next, the demand curve will look like the smooth one in Figure 9.3, rather than the stair-step demand curve of Figure 9.2. In Figure 9.3, the vertical axis shows the relative wage of women, w_f/w_m, rather than just women's wages. The horizontal axis shows the number of women. When $w_m = w_f$, there is some demand for women workers, stemming from firms for whom $d_j = 0$. These employers are willing to hire women even when their wages are the same as men's wages. That demand is the horizontal portion of the demand curve at $w_f/w_m = 1$. Most employers, however, are still unwilling to hire women at that wage. But, as w_f falls further and further, more

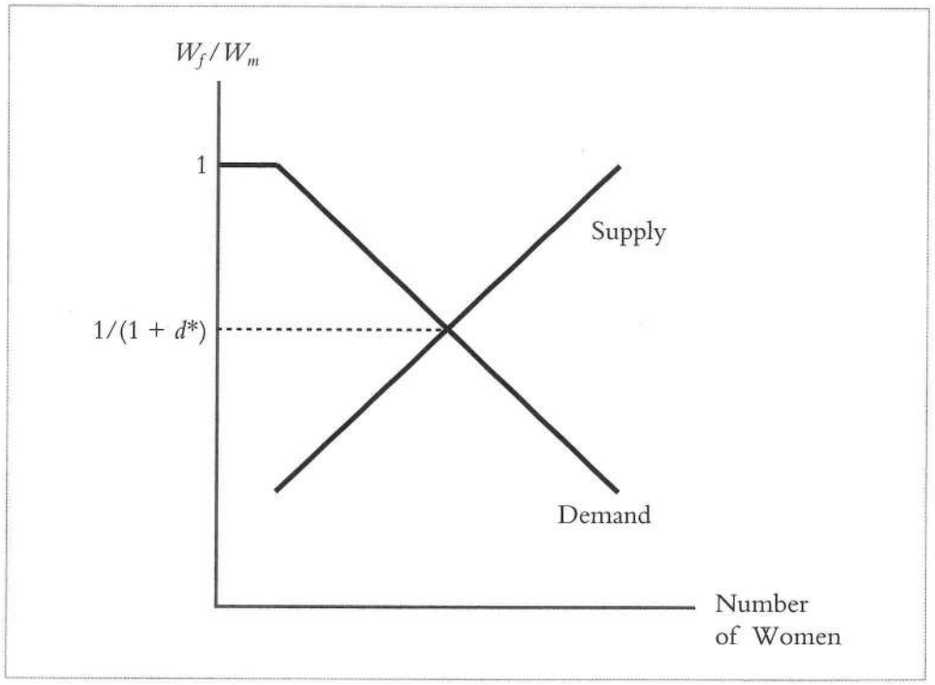


FIGURE 9.3 Women's Wages When Employers' Discrimination Coefficients Differ

and more employers are willing to hire women. Each one is just willing to hire women when $w_f \times (1 + d_j) = w_m$. Thus, the number of women workers demanded increases as the relative wage of women falls. The market equilibrium will be a relative wage at which the number of women willing to work just equals the number demanded. This is shown in Figure 9.3 as $1/(1 + d^*)$, where d^* is the value of d that just balances supply and demand. Women clearly continue to earn less than men. Again, it is clear that the relative wage will depend on both supply and demand; the greater the supply of women, the more prejudiced the employers they must deal with, and the lower their wages must be.

The equilibrium relative wage divides firms into two categories. For all firms with $d_j < d^*$, women are now relatively inexpensive compared to men because their own personal discrimination coefficient is low. For them, $w_f^* \times (1 + d_j) < w_m^*$, so they will employ only women and no men at all. They will pay them the market wage for women, which is lower than that for men.[14] For all firms with $d_j > d^*$, women are still too expensive relative to men because their own personal discrimination coefficient is so high. For them, $w_f^* \times (1 + d_j) > w_m^*$, so they will employ only men and no women at all. Only the marginal firm, for whom $w_f^* \times (1 + d_j) = w_m^*$, will be willing to hire both men and women.

Because of the difference in which kind of workers they hire, the profits of these two kinds of firms will differ, too. The firms hiring women have lower monetary labor costs than the firms hiring only men. Because the men and women are equally productive, the profits of the men-only firms must be lower than the profits of the women-only firms. It is easy to see why this must be true. The men-only firms could earn exactly the same revenues and have lower costs simply by replacing all their male employees with women because the women are just as productive, but cost less. That is a very powerful and important idea: *The more prejudiced firms sacrifice potential profits to indulge their prejudices. Discrimination is inconsistent with profit maximization.* It costs the firm something to indulge its discriminatory preferences.

There is one final step to the argument. Figure 9.3 might represent the state of affairs given the existing firms in the market. But the composition of the market might well change over time in a particular way. Potential firms with relatively low d_js could enter the market to take advantage, as it were, of the lower wages of women. Precisely because they are willing to employ women, they can earn the higher profits of the women-only firms in the market. But no firms with high d_js would want to enter the market because they would face the higher costs associated with hiring men.

[14]This is a subtle idea. Even firms with low discrimination coefficients pay women less than men because they take wages as given. These firms are *willing* to pay women more than that, but they don't have to.

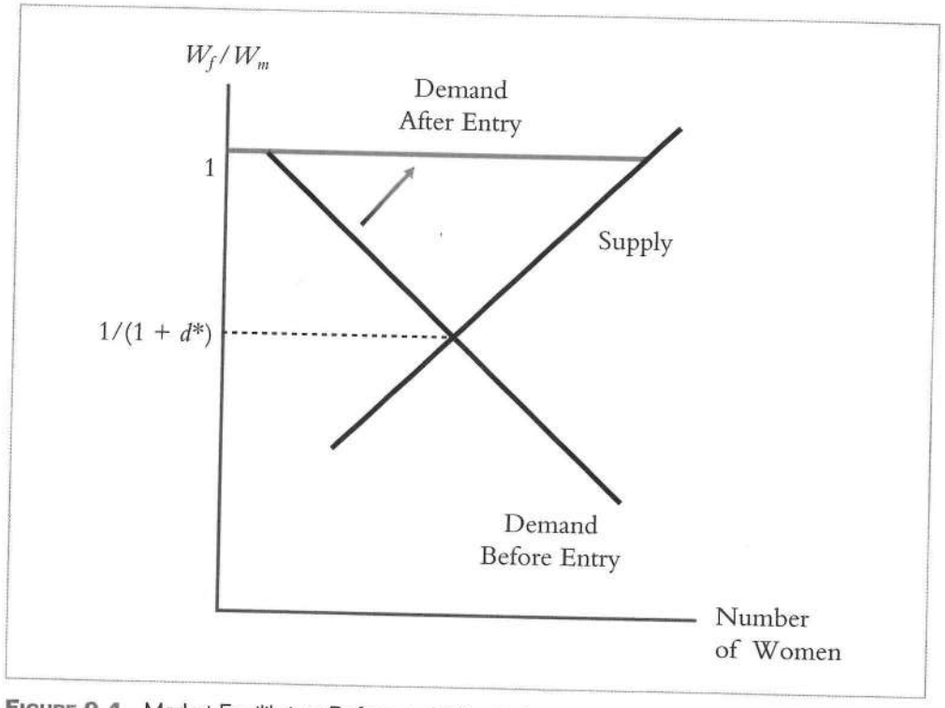


FIGURE 9.4 Market Equilibrium Before and After Entry

Figure 9.4 shows how this entry would change the market. The entry by these firms would shift the demand curve out at higher relative wages, thereby raising the equilibrium wage for women. Eventually, the entry process might result in a demand curve composed only of firms with $d_j = 0$. In that case, the demand curve would be horizontal at $w_m = w_f$. Discriminatory wage differences have been eliminated by competitive market processes! In a famous article, the economic theorist and Nobel Prize winner Kenneth Arrow showed that this result would be attained if there was free entry, constant returns to scale,[15] and a sufficient number of potential employers with $d_j = 0$. Needless to say, this is a surprising result and an important one. As Arrow wrote: "The model predicts the absence of the phenomenon it was designed to explain."[16]

It is important to appreciate the key elements necessary for the result—the ability to enter the market, the exclusive focus on monetary profits, and

[15]The constant returns to scale means that the market is likely to be competitive rather than dominated by a single monopolist or several oligopolists.

[16]Kenneth Arrow, "Some Mathematical Models of Race Discrimination in the Labor Market," in *Racial Discrimination in Economic Life* Anthony H. Pascal (editor) (Lexington, MA: Lexington Books, 1972). Arrow was one of the first U.S. economists to receive the Nobel Prize in Economics for his work on general equilibrium theory.

the existence of a sufficient pool of potential employers with low or even zero discrimination coefficients. Many economists, perhaps especially conservative economists, take the results largely at face value and conclude that earnings differences between men and women or, for that matter, any other two groups cannot reflect labor market discrimination but must, instead, be the result of genuine differences in productivity. (Those productivity differences could, of course, be caused by discrimination, perhaps public discrimination in the provision of public education.) The market will inevitably eliminate discrimination, they argue, by virtue of the profit-seeking activities of firms. For an interesting example of this process in the world of professional sports, see Box 9.2.

Box 9.2 Discrimination in Professional Sports

In a somewhat different context, discrimination in sports provides an interesting case exactly on point for both sides of the argument. For many years, discrimination in Major League Baseball was essentially absolute. No Black players at all were employed. There was no relative wage at which firms were willing to hire them. Instead, an alternative league, the Negro Baseball League, was formed to provide them an opportunity to play. They played there, often with great skill and always with very low wages. This situation prevailed for many years.

Why didn't the market just eliminate the unjustified earnings differences? One critical element of the "competition eliminates discrimination" argument was missing. Entry into Major League Baseball was not open. Would-be nondiscriminators could not simply set up shop and join the league, so race discrimination in baseball persisted for many years.

Eventually, however, the Brooklyn Dodgers, under the influence of a particularly broad-minded and cagey executive, Branch Rickey, signed Jackie Robinson to a contract as the first Black major league baseball player. He acted very much like an employer with a low value of d and recognized a great player—and a great bargain—when he saw it. The other teams acted like employers with high values of d. Within a decade, however, the color-bar had been broken on every team as a result of competitive pressures.

A similar story can be told for college athletics, where through the 1950s, most Southern schools fielded all-White football and basketball teams. Black athletes were not offered athletic scholarships. But this practice, too, eventually ended. When change finally came, it moved quickly across all universities.

The moral of this story is that discriminatory wage differences may persist for many years if the circumstances are right. But change, when it does come, can occur suddenly and thoroughly.

Other economists worry about the underlying conditions. Is entry sufficiently easy? Are firms sufficiently profit-oriented, or will they adhere to what may be social custom? Are there a sufficient number of unprejudiced employers? In this regard, the well-known economist and feminist, Barbara Bergmann,[17] quoted the yet more famous nineteenth-century economist, John Stuart Mill. Mill wrote that "Political economists ... exaggerate the effect of competition. ... They are apt to express themselves as if they thought that competition actually does, in all cases, whatever it can be shown to be the tendency of competition to do."

CUSTOMER DISCRIMINATION

The analysis of customer discrimination is quite similar to the analysis of employer discrimination. Now imagine that there are a number of firms selling an identical product in a competitive market. The firms are distinguished only by the gender, race, or other personal characteristic of the seller; the firms differ in no other important respect. If customers have no taste for discrimination, then the prices charged by the firms must be identical in long-run equilibrium. No customer would ever pay more for the product from one seller if it were available at a lower price from another seller. That would eliminate price differences. The "law of one price" would hold.

Suppose, though, that m (male) customers have tastes for discrimination. Now, faced with a good sold by an f (female) firm, these customers would regard the gross price as $p_f \times (1 + d)$, where d is, once again, the discrimination coefficient. The gross price of the product is composed of two parts—the monetary cost (p_f) and the disutility of interacting with an f firm ($d \times p_f$). If the f firm wants to sell its output to these customers, it must lower its price so that its gross price just equals the price of the good sold by m (male) firms. Thus, it must be the case that $p_f^* \times (1 + d) = p_m^*$. The discriminatory preferences result in f firms being forced to charge a lower price.

In thinking about how this might affect wage rates, imagine that the good in question is a service of some kind, where the customer interacts directly with the seller. For example, this could apply to doctors, dentists, lawyers, shop keepers, and all kinds of retail workers. Then, the price that is affected is either the worker's pay directly (as in the case of doctors, dentists, and lawyers) or the price of the good that the worker sells, which affects her pay indirectly.

If tastes for discrimination vary across customers, the analysis would follow exactly along the lines of the model of employer discrimination where the

[17]Barbara Bergmann, "Does the Market for Women's Labor Need Fixing?" *Journal of Economic Perspectives, 3* (1) (1989): 43–60.

discrimination coefficient varied. The f (female) customers would presumably have $d = 0$. Each male customer would have a critical value for p_f^*/p_m^* that was just low enough to make him indifferent between the two kinds of firms by making the gross prices equal. Market equilibrium would divide customers into those with lower discrimination coefficients who buy only from the f firms and those with higher discrimination coefficients who buy only from the m firms.

As in the case of employer discrimination, the customers who are indulging their taste for discrimination pay something for exercising those preferences. They pay a higher monetary cost by buying the good from m firms rather than f firms. They are not, however, worse off in terms of their own preferences—they quite willingly pay the higher price because they prefer to do so. Unlike the case of employer discrimination where firms with high discrimination coefficients earned lower profits and could potentially be driven out of business as a result, there is no corresponding mechanism here. Nothing happens to customers who prefer to buy goods from one firm rather than another except that they are worse off than they could be if they had different (no taste for discrimination) preferences. Consequently, customer discrimination can survive in market equilibrium in the long run.

EMPLOYEE DISCRIMINATION

If there is employee discrimination, then employees have tastes for discrimination concerning their co-workers. Suppose that m and f employees are again perfect substitutes. Employers have no taste for discrimination, but m employees don't like to work alongside f employees. Then, if the m money wage is w_m, their gross wage would be $w_m \times (1 - d)$ if they were working alongside women employees. Here, d enters negatively, reducing the gross wage received by m workers.

Now suppose there were two jobs that m workers could take, one with only m co-workers and one with both m and f co-workers. No m worker would take the latter job unless its money wage was sufficiently higher to offset the disutility of working with f workers. In symbols, if job 1 has only m workers and job 2 has m and f workers, then wages must be set so that $w_{1m} = w_{2m} \times (1 - d)$. The m wages in job 2 just compensate for the disutility of working alongside f workers. Women in this simple story are assumed not to care about their co-workers, so $w_{1f} = w_{2f}$. The apparent result is that men and women in job 1 will receive the same wage, but m workers in job 2 will earn more.

In fact, as Becker pointed out, an alternative result is that the workforce would be segregated by gender, thereby eliminating the need to pay higher wages to the men. Clearly, this is a lower-cost solution to the problem. So this simple model actually predicts job segregation, rather than pay differences.

In more complicated production settings, it is possible to establish wage differences on the basis of employee discrimination. For example, suppose there are m and f type 1 workers who are perfect substitutes for one another, but also m type 2 workers who are supervisors of the type 1 workers. Production requires either m or f type 1 workers and type 2 workers. If there is employee discrimination among type 1 m workers, job segregation will result, but, as above, there will be no pay differences. If type 2 m workers also have a taste for discrimination, they will receive higher pay when they supervise f workers than when they supervise m workers. In this situation, pay differences could arise from employee discrimination.

SUMMARY OF DISCRIMINATION MODELS

Let's review these three models of labor market discrimination. Employer discrimination can certainly affect wage rates in the short run—and the short run can last a very long time. Becker's insight that discrimination is costly to those who practice it is important, as is the prediction that market forces could operate to eliminate much of the earnings differences that are a result of discrimination. It is sensible to appreciate the insight but to be skeptical that it applies perfectly to labor markets in the twenty-first century.

Customer discrimination can certainly generate differences in prices for identical goods sold by different kinds of firms as long as there are tastes for discrimination. Unlike employer discrimination, this source of wage differentials will not be eroded by market forces. This model potentially applies primarily to services and retail where face-to-face contact is present. Employee discrimination probably has the least impact of all. Firms can readily avoid paying higher wages by segregating their workforce along the dimension on which discrimination operates.

Becker's models of labor market discrimination have been very influential among economists. Many, although certainly not all, economists interpret them as indicating that labor market discrimination is not likely to be an important factor in explaining the lower wages of women. We will look at the evidence in the next chapter.

Alternative Explanations

STATISTICAL DISCRIMINATION

Statistical discrimination is another form of discrimination that can arise in the labor market. It arises because employers have an interest in maximizing profits, which requires keeping costs as low as possible. Firms often bear the cost of hiring and training workers, and so it is in their interest to keep these

costs low. Unfortunately, the firm does not always have exactly the information it needs to determine which employees are likely to be the easiest to train, the most productive, the most reliable, or the most committed to the firm. This is called **asymmetric information**—the employee has more information than the employer about his/her own skills and commitment.

Because of this, the employer will often rely on a signal or a credential that the employee possesses to make a determination about whether to hire that person. For example, if graduates of Ivy League colleges are, in the firm's mind and experience, more productive than other college graduates, the firm may make its hiring decision based upon this criterion. The firm is using characteristics of a group (Ivy League college graduates) to make decisions about individuals and their productivity. The firm is following this course because it has imperfect information; it can't be certain about each individual's productivity, but it has found that, on average, people from this group are more productive.

From the perspective of women in the economy, this same kind of thing can happen if firms are looking to hire people into jobs where there are considerable training costs. The employer may believe that a particular woman is qualified for the job but may worry that she won't make the long-term commitment necessary to make the firm's investment in training her profitable. Perhaps in the past, the firm has had many women employees who left to raise a family. Because it is costly to investigate whether a particular applicant will leave the firm after a short period of time—and difficult to verify in any case—the firm may tend to offer jobs to men because they have historically had a stronger attachment to the labor force. Thus, women may find it more difficult to get hired at this firm. It isn't because employers actively dislike female applicants; it is because in the past female employees have tended to leave the firm, and the firm has had to incur the costs associated with hiring and training a new employee.

Figure 9.5 makes the point quite nicely. Suppose some attribute, call it attachment to the workforce, is distributed normally throughout the population but is higher for men than for women. We can draw two bell curves (normal curves) in Figure 9.5 to illustrate this distribution. These curves are drawn under the assumption that there exists a distribution of labor force attachment across men and women and that the distribution is different for each group—in particular, men have a higher degree of labor force attachment, and the variation of men's attachment is smaller than women's attachment. Think of the situation that existed thirty or forty years ago. In Figure 9.5 the average woman has a lower average attachment to the workforce (W) than the average man (M). However, the distributions do overlap a little bit: Some women are more attached to the workforce than some men. Employers

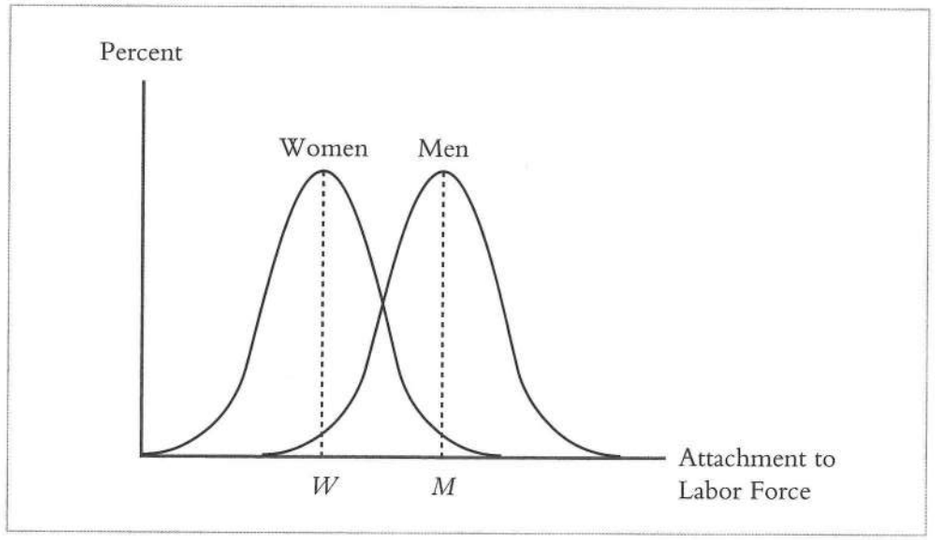


FIGURE 9.5 Statistical Discrimination Based on Different Distributions of Labor Force Attachment for Men and Women

who are looking for someone more attached to the workforce will likely select a man, given that they cannot observe a given woman's attachment to the workforce and that they know that men, on average, have been more attached to the workforce.

One might think that, given the strong attachment that women have demonstrated to the labor force over the past thirty years, statistical discrimination would be less likely to occur; after all, as we have documented throughout this book, men and women are becoming more alike in their attachment to the labor force. However, statistical discrimination can occur even if men and women have a nearly equal attachment to the labor force. This may happen if employers are risk averse and fear making a costly mistake or if the indicator of attachment is a noisy signal—it doesn't predict attachment well. Then, even though they understand that women are equally committed to the labor force, they may still prefer to hire men to minimize the costs associated with hiring someone who may leave the firm after a short while. Figure 9.6 illustrates this scenario with another set of bell curves. This time the distributions overlap nearly completely, indicating the near parity in labor force attachment that we now see. Here, some risk-averse employers could end up hiring a man who is less attached to the workforce than a woman.

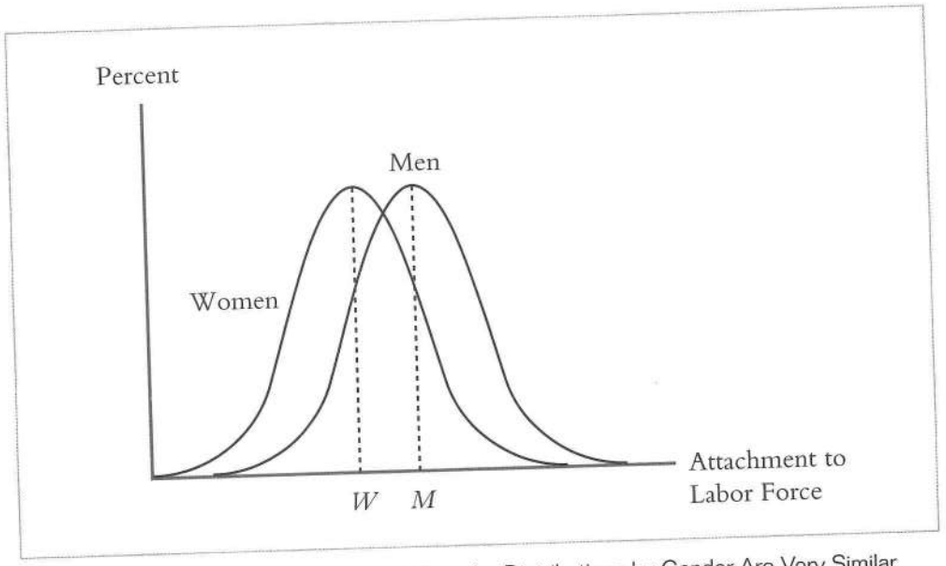


FIGURE 9.6 Statistical Discrimination When the Distributions by Gender Are Very Similar

Box 9.3 Women, Men, and Competition: Is This Why Women Don't Seem to Get Ahead?

The standard explanations of the gender gaps in the labor market focus on human capital differences, discrimination, and personal choices. Recently, experimental economists have proposed another potential explanation, based on laboratory experiments: Men perform better in competitive workplace environments.

As we have noted often throughout this text, economists regularly lament the difficulty of finding experimental data with random assignment and the lack of that data makes finding causal relationships between variables problematic. But sometimes economists can create their own data in the laboratory, filtering out all other complicating factors. In a paper published in 2003, economists Uri Gneezy, Muriel Niederle, and Aldo Rustichini ran an experiment in which groups of three men and three women solved computerized mazes in different kinds of "competitive" environments. In one experiment, individuals were rewarded based on the their own performance, a reward system called *piecework*. In this environment run, the men and women performed equally well. No statistically different gender differences in ability were observed.

Then the experiment was changed to what economists call a *tournament* reward structure, in which payoffs depend on an individual's relative score, just like in a golf or tennis tournament. In this experiment, the individual with the highest score was paid proportional to the number of puzzles he or she solved, while the others received nothing. Labor economists have noticed that many corporate pay structures have a tournament structure, with many "players" at the entry level, but only a few winners who make it to vice president and president of the company. Here, pay depends on relative, not absolute performance—the "best player" wins and gets the best salary, the second best becomes vice president and gets the second highest salary, and so on.

What happened in this experiment? The absolute performance of the women was the same as in the piecework experiment, but the absolute performance of the men improved. The competitive environment of the second experiment somehow elicited better performance from the men.

The experiments considered one more wrinkle. In the two experiments above, men and women competed against each other in mixed-sex groups. Some research suggests that women perform better in single-sex environments, either because they are less competitive in general or because they do not like competing against men. To test this, the experiments were also conducted with same-sex groups, first with piecework and then with a tournament pay structure. Gneezy, Niederle, and Rustichini found that, indeed, the women did perform better in the tournament structure than the piecework, unlike their performance in the mixed-sex environment.

In a related paper published in 2007, Niederle and Vesterlund examined whether men and women had different preferences for competitive environments. The general structure of the experiment was similar to the previous one—men and women perform tasks, in this case, adding up sets of five two-digit numbers for five minutes—under piecework and tournament pay systems. In this experiment, however, another issue was introduced. After performing the addition task in both pay environments, each individual was told his or her absolute performance level—how many sums were added correctly—but not how he or she did relative to others in the group. Then each was asked which compensation scheme they preferred for their next task.

No absolute differences in ability level existed between the men and the women. Both groups, on average, solved the same number of problems in the piecework environment and in the tournament system. Nevertheless, more than twice as many men as women chose the tournament pay scheme (73% versus 35%). Even low-ability men (men who solved fewer problems than average) opted for the tournament scheme, while even high-ability women opted for the piecework system. The authors conclude that men tend to be overconfident about their abilities.

These experiments and the two articles have many more rich insights than we have room to describe here. You may want to follow them up. The articles are accessible to a serious undergraduate student.

It's not an easy matter to know exactly what these experimental results mean for the differences between men and women in the real world of the labor market. Like many of the differences between the sexes, the findings are certainly intriguing. Big deal or not a big deal? And where do the differences come from in the first place? It's too early to tell.

Sources: Uri Gneezy, Muriel Niederle, Aldo Rustichini. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics*, *CXVIII* (August 2003): 1049–1074.

Muriel Niederle and Lise Vesterlund. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics 122* (3): (August 2007): 1067–1101.

EXPLAINING OCCUPATIONAL DIFFERENCES

As we explained in Chapter 8, **occupational segregation** by gender is still substantial, although it is on the decrease. The intriguing question is why it exists. This is potentially a very important question: While earnings differences by gender do exist within occupations (see Table 8.3), the difference by gender in the occupational distribution is an important part of the explanation of women's lower earnings. After all, it is not just that women and men tend to congregate in different occupations; it is that women tend to congregate in occupations that pay less than those in which men dominate. A full understanding of the cause of gender differences in occupation is probably not something we can realistically expect. But we do have a number of interesting possible explanations. As with most complex phenomena, each may be partly correct.

Gender Roles and Occupational Choice It is certainly possible that men and women truly prefer different kinds of occupations, as a result either of innate preferences or ubiquitous social cues about what is feminine and what is masculine. Studies of gender roles find that many individuals have definite ideas regarding what jobs or careers are suitable for what gender. These studies suggest that children as young as 4 years old show strong predilections toward occupations that are conventionally linked to a specific gender. Where these ideas of suitable occupations and roles for a specific gender came from is arguable. One idea is that jobs considered typically "masculine" demand traits considered typically masculine, such as aggression, competitive conduct, and physical strength. Jobs considered typically female

entail such traits as caring, compassion, and patience. These traits are considered to be innate under this theory and each gender "self-selects" occupations because of these differences.

Another view explains occupational segregation as a direct effect of the socialization process experienced while growing up. Children are exposed to what roles are appropriate for their gender and are thus taught what occupations are "appropriate" for them as adults. An adult who chooses a job that is not within widely accepted gender roles can endure mental stress and be ostracized by family, peers, and co-workers for his or her choice. Studies have shown that this is, in fact, the case. Women working in top management positions experience not just more mental stress than those in blue-collar jobs experience but also sexual harassment and even hostility from male co-workers. Though attitudes regarding the appropriate occupational choice for each gender have changed and converged to some degree, these traditional ideas still play an important role in selecting an occupation.

Whatever the underlying cause, it does not, however, follow immediately from this that women's occupations will necessarily have lower wages. For that to be so, one of two things must be true. First, it could be that women's jobs are more attractive than men's jobs in terms of their *nonpecuniary job characteristics*, holding the level of necessary skill constant. **Nonpecuniary job characteristics** are everything about a job except its pay—its working conditions, hours, safety, prestige, and so on. The theory of **compensating differences** tells us that when two jobs require the same skills, but one is widely held to have more desirable nonpecuniary job characteristics, then that job will have a lower equilibrium wage. It is not difficult to find examples that seem to reflect this.

Second, it could be that women are "limited" to a relatively small subset of occupations and thereby increase the supply of workers there relative to a world in which no gender stereotyping occurs. If supply is larger, then wages will be lower; this is simple supply and demand. That is the essence of the *crowding hypothesis.* This hypothesis has a logical internal consistency—if women are crowded into occupations, wages probably would be lower—but economists tend to be skeptical of models in which individuals are "compelled" to do things without clear ways in which the compulsion is enforced.

Human Capital and Occupational Choice More sophisticated models of occupational segregation focus on the characteristics of the occupations themselves in terms of human capital investment. One version is based on the theory of optimal investment in on-the-job training that we discussed earlier in this chapter (pp. 291–293). Investment in training is costly to an individual, usually in terms of lower earnings than could be earned in a job without training. Earnings are higher after the training is received, and thus the

investment has an internal rate of return. How high that rate of return is depends largely on how long an individual expects to stay in that job (for specific training) or in the labor market (for general training). It also depends on whether the individual is paid according to his or her productivity—that is, it depends on the absence of labor market discrimination.

On-the-job training typically is an inherent part of a job—if you have the job, you receive the training. The only way to avoid the training is to take a different job. Different occupations provide different amounts of training, so choice of the amount of training implies a choice of an occupation. Thought of this way, women might choose those occupations that offer the desired amount of on-the-job training. In a time period when women's work was irregular, congregating in occupations that provided relatively little training might make economic sense. The changing occupational distribution of the past decades also makes sense viewed in this light: As women's lifetime labor force participation increased, they upgraded their occupational choices and invested in more on-the-job training.

During the 1970s and 1980s, Solomon Polacheck took this argument a step further by postulating that occupations had different inherent penalties to intermittent labor force participation. In some, the impact is substantial—an individual, male or female, who leaves a job and returns to it some years later will have suffered a large wage penalty. In others, the impact is small, so that an individual could more or less pick up where he or she left off. The reasons for the difference in penalty could reflect the changing technological nature of a job—think of a computer programmer returning to work after a long absence. It could also reflect other peculiar institutional details of a job. For example, some jobs, such as nursing, have relatively low wage premiums to seniority, so the penalty is smaller.

Figure 9.7 shows how wage penalties might work. To make it simple, we focus on just two occupations, *j* and *k*. Assume that the skill level involved in the two occupations is relatively similar (at least well in advance) so that it makes sense to think of an individual potentially choosing at some earlier point in time between working in occupation *j* and working in occupation *k*. For each occupation, we've drawn two age-earnings profiles, one if the individual worked continuously and another if the individual dropped out of the labor market for a fixed interval of, say, five years. In Figures 9.7(A) and (B), the line *AB* is the likely wage path with continuous participation, and *ACDE* is the likely wage path if an individual is out of the labor force for the time period denoted by the two dashed lines. The wage penalty is shown in each figure as *DD'*.

Which job is a better choice depends on the profile selected. Occupation *k* has a higher wage, but a more substantial penalty, than occupation *j*.

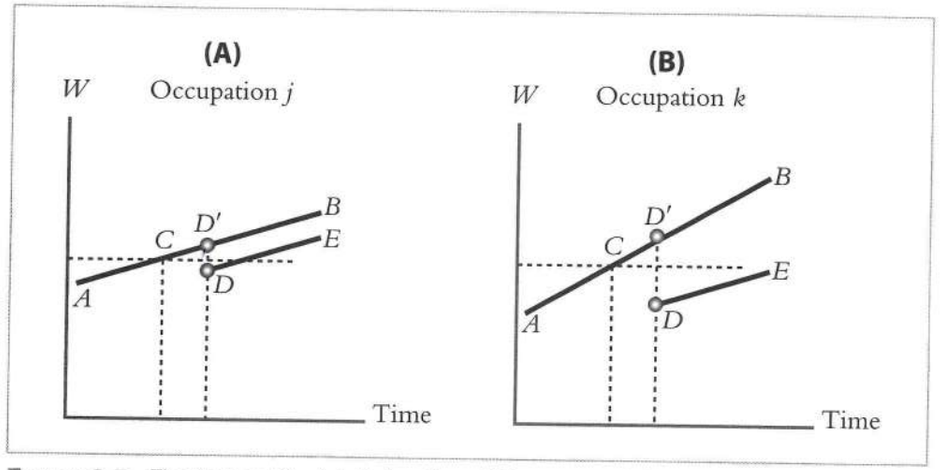


FIGURE 9.7 The Human Capital Explanation of Occupational Segregation by Gender—the Role of Wage Penalties to Periods of Nonwork

Clearly, for the individual who will work continuously, occupation *k* is the better choice. Technically, it has a higher present value of earnings. However, if the worker plans to leave the labor market for a number of years, occupation *j* is probably better, because he or she suffers a far smaller penalty. Thus, the two occupations offer individuals a tradeoff between wage levels and the penalty they will bear if they drop out of and then reenter the labor force. Individuals who expected to work continuously would probably congregate in occupation *k* and others like it, while individuals who expected to work intermittently would congregate in occupation *j* and others like it.

It's a small step to then argue that these factors could be part of the explanation of occupational segregation by gender. Women disproportionately end up—by choice in Polachek's model—in occupations like *j*, while men disproportionately choose occupations like *k*. Polachek's key point, then, is that women, knowing they will likely go in and out of the labor market, pick jobs that penalize them less for doing so. For example, they would choose to be a school teacher rather than a college professor, a clerk rather than a technician, a nurse rather than a doctor .

Needless to say, this argument is controversial. There probably is *something* to it. Teaching and nursing, two occupations with an obvious overrepresentation of women, are also occupations with a relatively small promotion ladder and thus not much of a penalty to lost years of experience. The empirical evidence on this explanation is mixed. The key relationship is the link between an occupation's penalty rate and its employment of workers with intermittent labor market

attachment. If Polachek is correct, there ought to be a negative relationship with workers with intermittent work patterns choosing the low-penalty occupations. Even though Polachek claimed that this was the case for wages and employment of men and women in eight occupational categories, Paula England argued that Polachek's model did not fit the facts quite as well as he claimed.[18] The details of their debate are reasonably technical, and ultimately, not particularly conclusive. Interested readers should consult the summer 1985 issue of the *Journal of Human Resources* to read the debate over this issue between England and Polachek. Perhaps the most interesting part of the debate is how angry the participants seem to be. This was a "hot-button" issue.

Summary

In this chapter, we've examined several major approaches for explaining why women earn less than men. Broadly put, earnings differences can exist either because the two groups of workers differ, on average, in their labor market skills, or because they are treated differently in the labor market, given the skills that they do have. The first of these is a human capital explanation; the second reflects labor market discrimination. We also more briefly examined statistical discrimination and theories of occupational segregation.

A human capital explanation of male-female earnings differences does have some resonance. We know that married households have often utilized a gender-based division of labor, with men specializing in labor market work and women in household production. A natural and inevitable consequence of that division is that, on average, married women spent far fewer years in the labor market than men did. This difference was quite large in earlier cohorts in which many women exited the labor market at marriage or at first birth and did not return, if at all, for many years. Differences in actual and/or expected labor force participation rationally lead to differences in investment in human capital, because the fewer years of participation lower the internal rate of return to an investment. It certainly doesn't help that the years of nonparticipation are those in which the present value of earnings is relatively high.

Becker's model of labor market discrimination does a great job of showing how prejudices could find their way into the demand for labor and could affect the relative wages that women, or other groups that are the victims of prejudice, earn. It also yields two important counterintuitive hypotheses: (1) The practice of discrimination is costly to the firms that practice it, and (2) strong competitive

[18]Paula England, "The Failure of Human Capital Theory to Explain Occupational Sex Segregation," *Journal of Human Resources*, 27 (1982): 358–70.

pressure might eliminate discriminatory firms from the market. Although this result is certainly not guaranteed, it is not an idea to ignore either.

When the human capital argument is combined with the "competition eliminates discrimination" argument, it leaves many more conservative economists believing that all earnings differences between groups in the labor market must be legitimate. In the next chapter, we use statistical analysis to evaluate whether the earnings gap between men and women is legitimate. We then turn to studies that provide real-world evaluation and computation of the sources of the gender gap in earnings.

Key Terms

asymmetric information 314
compensating differences 319
customer discrimination 303
discrimination coefficient 304
employee discrimination 303
employer discrimination 303
general training 290
human capital 288
internal rate of return 292
investments in human capital 289
non-pecuniary job characteristics 319
occupational segregation 318
on-the-job training 290
perfect substitutes 304
present value 291
specific training 290
statistical discrimination 313
taste for discrimination 303

Questions

1. Why is it necessary to discount future earnings to present value? What mistake do you make if you fail to do that?
2. Why is absence from the labor market at younger ages, for example, for raising a family, more important from a present value perspective than early retirement? Explain.
3. Use Excel to compute an internal rate of return to an investment with costs of $2,000 and benefits of $250 per year for 15 years. What happens to r^* if benefits in years 5 to 10 are $0 or if benefits in years 10 to 15 are $0?
4. Internal rate of return calculations are based on the historical experience of younger and older workers. Suppose today's young workers will not have the same higher earnings when they are older than today's older workers have. How would the corrected internal rate of return change? Would it be higher or lower?