Abstract

Applicants for any given job are more or less suited to fill it, and the firm will select the best among them. Increasing the wage offer attracts more applicants and makes it possible to raise the hiring standard, thereby improving the productivity of the staff. Wages that optimize on the trade-off between the wage level and the productivity of the workforce are known as selection wages. As men react more strongly to wage differentials than females, the trade-off is more pronounced for men and a profit-maximizing firm will offer a higher wage for men than for women in equilibrium.

The argument is not confined to issues of sex discrimination; rather it is of relevance for all labor markets where labor heterogeneity is important and supply elasticities vary systematically across occupations.

Keywords: Discrimination, sex discrimination, occupational discrimination, regional discrimination, selection wages, efficiency wages, hiring standards, monopsony, employment criteria, wage posting, Reder competition, wage structure, inter-industry wage structure, employer size-wage effect, occupational wage structure

Journal of Economic Literature Classification: J31, J7, B54, D13, D42
Introduction

Labor is heterogeneous. Applicants for any given job are differently suited to fill it. Faced with a number of applicants, a firm selects the best among them. Increasing the wage offer attracts more applicants and makes it possible to raise the hiring standard, thereby improving the productivity of the staff. Wages that optimize on the trade-off between the wage level and the productivity of the workforce are known as selection wages – a variant of efficiency wages (Schlicht 2005, 2007). The mechanism is explained in Section 1.

If the trade-off between the wage level and labor productivity differs for men and women, different wages for males and females will result and group discrimination obtains. This is to be expected. Empirical studies suggest, broadly speaking, that women and men differ in responsiveness to wage changes: the number of female applicants for a job opening reacts less to wage changes than that of males (Manning 2005, Ch. 7; Sulis 2007; Hirsch et al. 2008; Ransom and Oaxaca 2008; Hirsch et al. 2009). Hence an increased wage offer for males will attract more applicants and will permit a more pronounced tightening of the hiring standard for males as compared to females. Such behavioral differences would render it worthwhile for a profit-maximizing firm to offer a lower wage for women than for men. Section 2 develops this idea.

The core behavioral assumption used here – a different responsiveness of males and females to wage differentials – accords with prevailing social roles and stereotypes encountered in many economies that tie women more closely to the home than men. Such stereotypes, shared by men an women alike, effectively reduce the labor market options of women and render them less responsive to wage differentials. According to the view proposed here, this induces discrimination. Such discrimination feeds back on social roles and stereotypes. The process is explained in Section 3. A brief discussion of quotas and laws that enforce equal pay follows in Section 4.

The source of discrimination – the behavioral impact of social roles and stereotypes – is sometimes portrayed as a matter of “tastes.” This seems misleading, as such parlance assimilates widely shared attitudes and convictions to private preferences that have no bearing on others. Yet if such “tastes” are envisioned as conditioned by social roles and expectations, prevailing discrimination theories
contribute to our understanding of the phenomenon. Section 5 comments on these contiguous issues.

The selection wage theory of discrimination outlined here is closely related to the monopsonistic theory of wage discrimination, going back to Florence (1931), Robinson (1969, 302-4), and Madden (1973, Ch. 3). Section 6 discusses the connection, and provides some remarks on empirical implications that are specific to the selection wage theory.

The argument is developed with a view on sex discrimination, but it is applicable whenever a selection problem arises, hiring standards are implemented, and supply behavior plays a role. Section 7 illustrates this by discussing a hypothetical example where workers who work in banking are more responsive to monetary incentives, as compared to workers in care – healthcare, childcare, etc. It is shown that such differences in supply behavior across occupations may induce wage differentials that are not related to scarcity, are socially inefficient, and certainly unfair – just another case of discrimination.

1 Selection Wages

Consider a labor market where workers differ in productivity. There are a number of firms operating in the market. Each of these firms employs a number of workers and offers a certain wage. In the aggregate, this results in a certain level of employment $N$ and a certain average market wage rate $W$.

Labor supply for a single firm is described by a function

$$f(\theta, w, W) > 0.$$ (1)

This function gives the number of workers of productivity $\theta$ willing to work at that firm for the wage $w$, given the market wage level $W$.\(^1\) The supply function is the

\(^1\) For simplicity of exposition, all variables are taken as real. Further, and for the same reason, a pure time rate is presupposed. In presence of performance pay, the same arguments would hold true as long as performance pay is not such that firms are indifferent between hiring workers of different productivity. This can be ruled out from an empirical perspective, as wage compression is substantial (Frazis and Loewenstein 2006; Frank 1984; Bishop 1987; Bewley 1999, 85) and firms, as a rule, actually do screen their applicants carefully. See also Frank (1984) and Schlicht (2007, 18-20) for theories on wage compression.
Figure 1: (a) The labor supply function $f(\theta, w, W)$ gives the number of workers of productivity $\theta$ willing to work at the firm for the wage $w$, given the market wage level $W$. Average productivity is $\bar{\theta}$.

(b) If the firm increases the wage offer $w$, supply increases for all grades of labor. A decrease in the market wage level $W$ would entail a similar change.

We assume that an increase of the wage offer $w$ increases the supply of all grades of labor. Denoting partial derivatives with subscripts, this is expressed formally by writing $f_w > 0$. The effect of wages on supply decreases with an increasing wage offer ($f_{ww} < 0$). For any given wage offer $w$, an increase in the market wage level $W$ decreases supply as this indicates an improvement in outside options ($f_W < 0$). A decrease of the market wage level $W$ makes any given wage offer $w$ comparatively more attractive and can be assumed to act on the marginal effect of a wage increase ($f_w$) like an increase of the wage ($f_{wW} > 0$).

Figure 1(b) illustrates the effect of wage changes on supply. Formally we have

$$\theta \in [\theta_0, \theta_1], \ f_w > 0, \ f_{ww} < 0, \ f_W < 0, \ f_{wW} > 0.$$  

Note that labor supply is to be conceived as comprising all workers that could be hired at the wage offer $w$ by the firm under consideration, irrespectively of whether they are employed elsewhere, or unemployed.
The firm has a technologically fixed number of jobs \( n \) to fill. The minimum wage that is required to fill these jobs is implicitly defined by

\[
    n = \int_{\theta_0}^{\theta_1} f(\theta, w_0(W), W) \, d\theta
\]

and is denoted by \( w_0(W) \). It is the competitive wage. Because of

\[
    w'_0 = -\frac{\int_{\theta_0}^{\theta_1} W f W \, d\theta}{\int_{\theta_0}^{\theta_1} f W \, d\theta} > 0
\]  

the competitive wage is an increasing function of the market wage level \( W \).

The firm has to pay at least the competitive wage \( w_0 \) in order to fill its vacancies, but it can offer more. If it does so, it will improve the number of applicants, including more productive applicants. This makes it possible to screen workers and select only the best applicants, that is, only applicants with productivity above a certain threshold value \( s(w, W, n) \) that is implicitly defined by the equation

\[
    n = \int_{s(w, W, n)}^{\theta_1} f(\theta, w, W) \, d\theta.
\]  

The threshold \( s(w, W, n) \) is the hiring standard. From (4) we obtain

\[
    s_w = \frac{1}{f} \int_{s}^{\theta_1} f W \, d\theta > 0, \quad s_W = \frac{1}{f} \int_{s}^{\theta_1} f W \, d\theta < 0, \quad s_n = -\frac{1}{f} < 0.
\]  

The firm can increase its hiring standard if it offers a higher wage rate \( s_w > 0 \). If the market wage level increases, supply at the given wage offer is reduced and the firm must lower its hiring standard \( s_W < 0 \). Further, the hiring standard is to be reduced if more workers are to be hired \( s_n < 0 \). Note that workers with productivity less than \( s \) that are not hired are not necessarily unemployed, as they may be employed elsewhere, although they would prefer employment at the firm under consideration, given its wage offer.
The overall wage rate $w$ and a market wage level $W$ imply for the firm average productivity

$$a(w, W, n) = \frac{s(w, W, n)}{\int_{s(w, W, n)}^{\theta_1} f(\theta, w, W) d\theta} \int_{s(w, W, n)}^{\theta_1} \theta^2 f(\theta, w, W) d\theta = \frac{1}{n} \int_{s(w, W, n)}^{\theta_1} \theta f(\theta, w, W) d\theta. \quad (6)$$

This is the productivity curve depicting the relationship between the wage offer and the productivity of the firm’s workforce (Figure 3). Its derivatives are

\[
\begin{align*}
  a_w &= \frac{1}{n} \int_{s(w, W, n)}^{\theta_1} (\theta - s) f_w d\theta > 0 \\
  a_{ww} &= \frac{1}{n} \int_{s(w, W, n)}^{\theta_1} (\theta - s) f_{ww} d\theta - \frac{1}{n f(\int_{s(w, W, n)}^{\theta_1} f_w d\theta)^2} < 0 \\
  a_W &= \frac{1}{n} \int_{s(w, W, n)}^{\theta_1} (\theta - s) f_W d\theta < 0 \\
  a_{wW} &= \frac{1}{n} \int_{s(w, W, n)}^{\theta_1} (\theta - s) f_{wW} d\theta - \frac{1}{n f(\int_{s(w, W, n)}^{\theta_1} f_w d\theta d\theta) (\int_{s(w, W, n)}^{\theta_1} f_w d\theta)} > 0 \\
  a_n &= -\frac{1}{n^2} \int_{s(w, W, n)}^{\theta_1} (\theta - s) f d\theta = -\frac{1}{n} (n - s) < 0 \\
\end{align*}
\]

where $s = s(w, W, n)$. In other words: average productivity increases if the wage offer $w$ increases. The effect diminishes with an increasing wage rate. An increase in the average wage level $W$ works similar to a decrease in the wage offer $w$, as it improves outside options of the workers. This implies the third and fourth inequalities in (7). Finally we note for later use that an increase in $n$ decreases average productivity.

The firm faces therefore a trade-off between the wage level and the productivity achievable. The argument is illustrated in Figure 2. The resulting productivity curve is depicted in Figure 3(a). According to (3) and (7) the competitive wage increases with an increasing market wage level $W$, and the productivity curve shifts down.

---

2 The concept of the productivity curve has been introduced in Schlicht (1978, 393).
Figure 2: (a) Employment \( n \) implies a hiring standard \( s \) and average productivity \( a \). Workers with productivities below \( s \) are not hired. (b) Increasing the wage offer increases supply and permits the firm to increase its hiring standard to \( s' \). This entails an increased average productivity \( a' \).

Figure 3: (a) The minimum wage the firm must pay is the competitive wage \( w_0 \). Increasing the wage above that level permits a tightening of the hiring standard, engendering increased productivity, as described by the productivity curve \( a(w, W, n) \). An increase in the wage level from \( W \) to \( W' \) shifts the productivity curve down to \( a(w, W', n) \) and increases the competitive wage from \( w_0 \) to \( w'_0 \). (c) The efficiency wage \( w^* \) is given by the condition that the slope of the productivity curve equals \( \frac{1}{p} \).
Given the product price \( p \), the firm will maximize the difference between the productivity of its staff, valued at the market price of the product, and labor costs by selecting the appropriate wage rate.\(^3\)

\[
p a (w, W, n) - w \rightarrow \max_w \quad (8)
\]

We assume that the product price is high enough to assure profitability. The first-order condition for a profit maximum is

\[
p a_w - 1 = 0. \quad (9)
\]

As \( a_{w w} < 0 \), the second order condition is satisfied, equation (9) may alternatively be written as

\[
\varepsilon (w, W, n) = \frac{w}{p a (w, W, n)} \quad (10)
\]

with

\[
\varepsilon (w, W, n) := \frac{\partial a_w}{\partial w}
\]

as the elasticity of productivity. Condition (10) states that the elasticity of productivity is to be equal to the ratio of labor costs to sales minus variable costs. This elasticity condition differs from the Solow condition usually encountered in efficiency wage theory which requires the elasticity of productivity to equal one in equilibrium (Solow 1979, 80; Schlicht 1978, 340). The difference is due to the different (and, it is hoped, more realistic) assumption regarding the technology of the firms.

Equation (9) implies an optimal wage rate

\[
w^* = w(W, p, n)
\]

which is referred to as the selection wage. Its determination is depicted in Figure 3(c). Regarding its properties we find

\[
\begin{align*}
w^*_W &= - \frac{a_{w w}}{a_{w w}} > 0 \\
w^*_p &= - \frac{a_w}{p a_{w w}} > 0.
\end{align*} \quad (11)
\]

\(^3\) The product price \( p \) is to be understood as the contribution margin of a worker with \( \theta = 1 \), that is, the receipt for the production produced by such a worker minus variable non-labor costs.
In other words: an increasing market wage will induce the firm to offer a higher wage rate, and an increasing product price will increase the wage offer, too.

2 Sex Discrimination

Now consider the case of two groups of workers, males and females. Superscripts $f$ and $m$ refer to females and males, respectively. Hence $f^f$ and $f^m$ denote the respective supply functions, $W^f$ and $W^m$ refer to the respective market wage levels, $w^f$ and $w^m$ stand for the firm’s wage offers, and so forth.

The firm has to fill $n$ jobs by hiring men and women:

$$n = n^f + n^m$$  \hfill (12)

With a hiring standard $s^f$ for women and a hiring standard $s^m$ for males, the average productivity of females is $a^f$ and that of males is $a^m$. The productivity of the work force is thus

$$a = \frac{n^f}{n} a^f(w^f, W^f, n^f) + \frac{n^m}{n} a^m(w^m, W^m, n^m).$$

The average wage rate is

$$w = \frac{n^f}{n} w^f + \frac{n^m}{n} w^m.$$

The firm wants to maximize profits $pa - w$ which is:

$$p \left( \frac{n^f}{n} a^f(w^f, W^f, n^f) + \frac{n^m}{n} a^m(w^m, W^m, n^m) \right) - \left( \frac{n^f}{n} w^f + \frac{n^m}{n} w^m \right)$$

with regard to $w^f$, $w^m$, $n^f$, and $n^m$ under constraint (12). This gives rise to the Lagrangian

$$\mathcal{L} = p \left( \frac{n^f}{n} a^f(w^f, W^f, n^f) + \frac{n^m}{n} a^m(w^m, W^m, n^m) \right) - \left( \frac{n^f}{n} w^f + \frac{n^m}{n} w^m \right) + \lambda \left( n - n^f - n^m \right).$$  \hfill (13)
The first-order conditions for a maximum are
\[ \frac{\partial L}{\partial w_i} = \frac{n^i}{n} (p a^i w_i - 1) = 0, \ i = f, m \tag{14} \]
\[ \frac{\partial L}{\partial n^i} = \frac{p}{n} (a^i + n^i a^i_n) - \frac{w_i}{n} - \lambda = 0, \ i = f, m \tag{15} \]
In view of (7) this implies
\[ a^f = a^m = \frac{1}{p} \tag{16} \]
\[ p \cdot s (w^f, W^f, n^f) - w^f = p \cdot s (w^m, W^m, n^m) - w^m \tag{17} \]

The first condition is similar to the condition (9). It states that the group-specific wage is determined by the condition that the slope of the productivity curve is equal to \( \frac{1}{p} \). Hence wages will be the same if the productivity curves are the same. The second condition says that the difference between the value of the hiring standard and the wage must be the same for both groups.

Consider the case that males and females have identical supply functions and the market wage levels \( W^f \) and \( W^m \) are the same. In this case, the symmetric solution \( w^f = w^m \) and \( s (w^f, W^f, n^f) = s (w^m, W^m, n^m) \) will satisfy (14) and (15). No discrimination would occur. It can readily be seen that this result carries over to the case where labor supply for females is just a fraction \( \alpha \) of the supply of males, amounting to \( f^f = \alpha f^m \) for all \( \theta, w, W \) and \( n^f = \alpha n^w \).

However, this equal treatment will be ruled out if supply behavior of females differs from that of males. It has been hypothesized and empirically supported that this may indeed be the case (Manning 2005, Ch. 7, Sulis 2007, Hirsch et al. 2008, Ransom and Oaxaca 2008, Hirsch et al. 2009). Women’s labor supply seems, on average, more elastic than that of males. Ransom and Oaxaca 2008 estimate elasticities of 1.5 for women and 2.7 for men, for example, when looking just at averages. This broad regularity may be formalized by assuming \( f^f_w (\theta, w, W) < f^m_w (\theta, w, W) \) for all \( \theta, w, W \) and \( a^f_w (w, W, n) < a^m_w (w, W, n) \).

**Theorem:** If, for some values \( (w, W) \), the following assumptions are satisfied, the Lagrange conditions (14) and (15) are violated:
\[ f^f (\theta, w, W) = f^m (\theta, w, W) \text{ for all } \theta \in [\theta_0, \theta_1] \]
\[ f^f_w (\theta, w, W) < f^m_w (\theta, w, W) \text{ for all } \theta \in [\theta_0, \theta_1]. \tag{18} \]
In other words: If the supply and the distribution of productivities of men and women is identical at some common wage offer \( w \), given some shared market wage rate \( W \), while supply of males reacts more strongly to deviations from the common wage, the individual firm will have an incentive to offer discriminatory wages.

Proof: If the market wage level \( W \) and the wage rate \( w \) are identical for men and women, condition (17) implies that the same hiring standard \( s \) is applied to men and women. Hence the same fraction of men and women are employed and their average productivities will be the same. As the women’s labor supply exhibits a lower sensitivity to wage changes \( (f^w_f < f^m_m) \), the average sensitivity of supply \( (a^f_w \), and \( a^m_w \), respectively) taken in both cases over the range of productivities \([s, \theta]\) must be less for women than for men. This violates (14).

\[ \text{QED.} \]

It can readily be seen that the theorem carries over to the case that labor supply for females is just a fraction \( \alpha \) of males, amounting to \( f^f = \alpha f^m \) and \( f^w_w < \alpha f^m_w \) for all \( \theta \) and some \( w, W \).

To elucidate matters further, consider the case where men and women are lumped together. This results in a supply function

\[ f(\theta, w, W) = f^f(\theta, w, W) + f^m(\theta, w, W) \]

and wage setting would proceed as in Section 1. We call this the pooling equilibrium. Assume that at this equilibrium (18) obtains, viz. that at the common wage rate \( w \) the forthcoming productivity distribution of males and females is exactly the same. Average productivity would be

\[ a = \frac{1}{2} a^f + \frac{1}{2} a^m \]

where employment levels for males and females \( n^f = n^m = \frac{1}{2} n \) are induced by the shared hiring standard \( s \). Regarding the sensitivity of productivity to wage changes, we would have (still with the same hiring standard and wage for men and women)

\[ a_w = \frac{1}{2} a^w_w + \frac{1}{2} a^m_w. \]

As \( a_w = \frac{1}{\alpha} \) from (9) and \( a^f_w < a^m_w \) from (18) we must have, in violation of (16),

\[ a^f_w < \frac{1}{\alpha} < a^m_w. \]
By lowering wages for women and increasing wages for men while keeping employment levels and the average wage unchanged, the firm can increase its profits: increasing the wage for males by $\Delta$ while keeping the employment of males and females unchanged requires a reduction of the wage for women by the same amount. Such a change in payments would leave the wage bill unchanged. Yet the induced change in average productivity is

$$\frac{1}{2} (a^m_w - a^f_w) \Delta > 0.$$  

Hence the firm can, by introducing a discriminatory wage policy, increase the productivity of its work force without changing the wage bill. Such a change goes along with tightening the hiring standard for males and loosening the hiring standard of females in order to accommodate the change.

The argument carries over to the case where supply of females is not identical to that of males, but just a fraction of the supply of males for all $\theta$.

Once the firm is offering different wages, it will have an incentive to adjust the share of women and men in total employment in order to meet condition (17).

### 3 Market Discrimination and the Social Multiplier

According to the view expounded here, wage discrimination according to sex can be triggered by differences in supply behavior of males and females, even if no initial discrimination exists at the market level. As the argument applies to all firms in the market, we would expect market wage levels for men and women to diverge. This will feed back on discrimination.

According to (11), an increased market wage level for one of the groups will push the firm’s optimal wage offer up, and thereby increase the firm’s wage offer even further, and so on. Any initial discrimination will be augmented at the market level. This process is known as the “social multiplier” (Schlicht, 1981; Glaeser et al., 2003 and, regarding discrimination, Schlicht, 1982). Applied to sex discrimination, the argument would be that firms offer slightly lower pay to women, and this will reduce the market wage level for women, inducing firms to reduce the wage offer for women even further. Due to this process, we would
expect significant discrimination to emerge at the market level even if initial wage differentials are rather small.

Such a process may be seen as a kind of self-confirming cycle, involving two mechanisms, one economic and the other psychological. The economic mechanism is the following. Assume a society with a traditional sexual division of labor that ties women more closely to the home than men. The entailed behavioral differences trigger wage discrimination. This economic discrimination supplies an incentive for maintaining the traditional pattern of a sexual division of labor, and thereby stabilizing discrimination. Given economic discrimination, psychological mechanisms stabilize and amplify supporting attitudes. Any persistent discrimination will appear as normality after a while, and re-establish the social norm. As Kahneman et al. (1986, 730f.) put it: “Psychological studies of adaptation suggest that any stable state of affairs tends to become accepted eventually, at least in the sense that alternatives to it no longer readily come to mind. Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction. Thus, the gap between the behavior that people consider fair and the behavior that they expect in the marketplace tends to be rather small.” Discrimination will appear “natural” and will become an integrated part of the identities of men and women, and these identities entail these “natural” behavioral differences that give rise to different supply behavior – possibly generalizing to behavioral patterns even of men and women without family responsibilities.

4 Fighting Discrimination

A possibility of combatting discrimination would be to introduce a quota system, but this would not eliminate discrimination. Consider the situation envisaged in the proof of the theorem: at a common wage rate, the skill distribution of both populations are identical, but women react less to wage changes than men do. As discussed above, this would create an incentive for firms to pay women less than men. In order to be effective, a quota system must be supported by further measures, such as an anti-discrimination law that enforces equal pay for men and women doing the same job.
Such an anti-discrimination law would by itself enforce identical hiring standards for both groups and render a quota system superfluous – see (17) above. Yet the incentive to discriminate would not be removed, and firms might actually implicitly discriminate in a system where explicit wage discrimination is prohibited by law. This could be done by creating formally different positions that are paid differently yet involve identical tasks. Another way to circumvent anti-discrimination regulations could be to promote women sluggishly, etc. All these problems may reduce the effectiveness such measures to some extent.

Note that the statistical theory of discrimination would require a quota system along with equal pay enforcement in order to avoid discrimination through segregation. The selection wage theory, as well as the closely related monopsony theory, differ here, as no quota system seems to be required.

A more radical approach to challenge sex discrimination – call it the “Scandinavian solution” – has been proposed by Eskil Wadensjö (2009). It aims at fighting discrimination at the root by implementing policies that assimilate the social roles of men and women, and associated stereotypes. This could eliminate, or at least reduce, differences in supply behavior, and discrimination would fade.

## 5 Wage Differentials and Social Stereotypes

Socially, wage discrimination is particularly objectionable if it is based on social prejudices and stereotypes that are unrelated to productivity differentials between the groups involved. (We do not, as a rule, object to more productive workers being better paid, unless pay differentials are glaringly out of proportion.) Hence a central tenet of any theory of labor market discrimination must be to establish a link between wage differentials and social stereotypes. The main theories (taste discrimination, statistical discrimination, and monopsonistic discrimination) all contribute to such an understanding, but have also some shortcomings. Consider this in somewhat more detail.

*The taste theory of discrimination.* This theory assumes fixed discriminatory preferences (Becker, 1971). We may conceive these preferences are engendered by given social norms and stereotypes. In this theory, it remains sometimes unclear in which way such stereotyped preferences may generate wage discrimination that
can persist in the marketplace. If discriminatory behavior is seen as brought about by a “taste for discrimination,” and if this is treated like any other taste, then those who want to satisfy it have to pay for that by giving up income. Hence we would expect groups with discriminatory preferences would earn less than groups that are indifferent in this regard, but it seems not to be typical that discriminators are worse off than the victims of discrimination.

In a more fundamental way, the parlance about a “taste” seems highly problematic, mainly for two reasons: The first is that discriminatory preferences are taken as given (Lundahl and Wadensjö, 1984, 89-91). Yet, as Alfred Marshall has observed in a related context, “a great influence is exerted by custom and public opinion which are themselves the outcome of the experience of past generations.”

In other words: The social norms that trigger discrimination are endogenous, they are not arbitrary, and cannot be taken as givens for purposes of long-term analysis. There is a feed-back from outcomes to norms.

The other objection to assimilating discriminatory norms to personal tastes pertains to the issue of externalities, “taste externalities,” if you will. While it does not matter to others if you prefer tea over coffee, it makes a difference for women and men what social norms (“tastes”) prevail in society. A wide-spread “taste” for discriminating women reduces the welfare of women. In this sense it is not a private affair. It may be objected that a preference for coffee will hurt the tea merchants. This is true, but the tea merchants may change their occupation, while sex is not a matter of choice. The “taste”-parlance tends to fade out the important ethical dimension.

The statistical theory of discrimination. This theory may explain why some indicators (like sex) convey probabilistic information about productivities and induce wage differentials (Phelps, 1972). If social roles or stereotypes exist that actually trigger different behaviors of the groups, the observation that an applicant is female may signal a higher probability of unfavorable events in the prospective employment relation, perhaps a higher probability of terminating the employment relationship prematurely for family reasons. If the firm plans to invest in the

---

4 Marshall (1920, vi.v.ix), emphasis added.
5 Older formulations of the taste theory were aware of the possibility that economic discrimination itself may trigger discriminatory attitudes, see Wooton (1946, 115) and Harrod (1946, 91) for example.
worker, such a prospect will make the firm reluctant to hire a woman unless a lower wage or better productivity signals compensate the drawback (Lazear and Rosen 1990, Sattinger 1998). As social roles are internalized by both males and females, corresponding behavioral differences are to be expected. In this way, such a theory ties social stereotypes to economic outcomes. Such discrimination is perfectly efficient, as the labels “male” or “female” signal that the workers differ in productivity, if only probabilistically. From an efficiency point of view this warrants different treatment. Hicks (1946, 99) expressed this as follows: “But if there is a general presumption that the abilities of men and women are likely to be appreciably different, so that the sex of a worker is some significant guide to his or her efficiency, then it would seem that some discrimination on a sex basis is conductive to the efficiency of industry and probably to economic welfare in general.” Yet if the term “discrimination” is construed as referring to differential treatment of workers of equal productivity, the statistical theory of discrimination does not provide a theory of discrimination in this sense. It treats wage differentials between men and women just as other wage differentials. Further, it faces difficulties in accounting for group discrimination (Cain, 1986, 724-9).

The monopsonistic theory of discrimination. This theory relates wage differentials not to productivity but to supply behavior. If supply behavior is shaped by social roles and stereotypes, the theory may account for forms of discrimination that do not relate to productivity but rather to social roles. In this spirit it has been suggested and empirically supported that discrimination may be traced to differential supply behavior of males and females (Florence 1931; Robinson 1969, 302-4; Madden 1973, Ch. 3; Schlicht 1982; Manning 2005, Ch. 7; Sulis 2007; Hirsch et al. 2008; Ransom and Oaxaca 2008; Hirsch et al. 2009; Hirsch 2009). Women place more emphasis on non-pecuniary attributes of the workplace than men do. They value proximity to home and agreeable working hours while males are more responsive to monetary incentives.  

\[6\] It may be conjectured that firms may use such non-monetary job features to attract women, and this may give rise to another mechanism. This idea is not pursued here in order to establish the argument that differential pay of equally productive workers is a possibility, even sometimes a necessity, in an unregulated labor market.
seen from the point of view of the firm, will be less elastic than that of males, and this gives rise to discrimination, including group discrimination.\footnote{The assumption that female labor supply, as seen from the point of view of the firm, is less elastic than that of males does not rule out that market supply for women may be more elastic than that of males. A brief statement can be found in the Appendix 7.}

\textit{The Selection Wage Theory.} This theory, proposed here, builds on the same behavioral assumption regarding supply behavior and develops a related idea in an efficiency-wage context.\footnote{The fundamental idea of drawing on efficiency wages in order to explain discrimination goes back to Stiglitz (1973). Schlicht (2001, Section 5) provides a variant in the turnover context that generates job discrimination.} The outcome is again triggered by behavioral differences entailed by social roles and stereotypes and leads to group discrimination.

The theory does not rule out that social norms (“a taste for discrimination”) may be important proximate causes for generating discrimination; rather it offers a perspective to explain the formation and persistence of such norms. Neither does the selection wage theory rule out that statistical productivity indicators are connected to social stereotypes. In this sense, the selection wage theory offers not an alternative, but a complement, to the existing approaches.

6 \hspace{1em} \textbf{Monopsonistic Wages and Selection Wages}

Selection wage theory complements existing approaches and provides, in a certain sense, an extension of monopsony theory. The empirical support for monopsonistic wages lends support to one part of the mechanism underlying the selection wage theory: a less than infinite supply elasticity for the various grades of labor, as seen by the firm. The selection problem arises in such a context if labor is heterogeneous, and hiring standards are used as policy instruments by the firms.

There is considerable evidence that firms actually use both wages and hiring standards as instruments of adjusting to varying conditions in labor markets. Devereux (2002) and Büttner et al. (2009) show that firms do engage in upgrading and downgrading of jobs over the business cycle, quite in line with Reder’s (1955) classical thesis. Further, the literature on the employer size-wage effect provides some arguments that support the idea that hiring standards are important instruments.
The employer size-wage effect relates to the observation that larger employers pay significantly higher wages than smaller ones, on the order of magnitude of about 20%, which is comparable to numbers reported about sex discrimination (Schwimmer, 2007). There is the further observation that larger firms put much more effort into selecting their workers than smaller firms do, and impose more demanding hiring standards. Guided by these findings, Schwimmer (2007) has contended that larger firms place more emphasis on the productivity of their workers because they implement a deeper division of labor than smaller firms. This finer division of labor enables them to achieve higher productivity, but entails a larger share of co-ordination costs. Because co-ordination costs are disproportionately more important for larger firms, the qualification of the work force is more important to them than to smaller firms. Smaller firms would, of course, prefer better workers, too, but not at the price the larger firms are willing to offer. In order to implement a more demanding hiring standard, larger firms offer better wages. This is precisely the selection aspect emphasized here: the observation that hiring standards are related to employer size, and that the number of applicants for a given job opening seems to be larger in large firms suggests that hiring standards do play a significant role in hiring decisions.

Although there is a strong overlap of arguments underlying the monopsonistic and the selection wage theories, there are also some differences that may help to empirically separate the effects of the two mechanisms. For monopsony, only supply elasticities count and hiring standards must be taken as given, while the selection wage theory also incorporates the influence of labor heterogeneity on wage setting. Bringing supply elasticities of men and women more in line would reduce discrimination in both cases, whereas an increase in the importance of individual productivity differentials within firms, brought about e.g. by switching

---

9 The employer size-wage effect has been attributed in part to monopsonistic wage setting (Burdett and Mortensen 1998; Manning, 2005, Ch. 4): Larger firms need more workers and have, therefore, to pay more. Yet the number of applicants per vacancy seems to be larger in large firms, which renders the position problematic, see (Brown and Medoff 1989, 1049; Gerlach and Hübler 1998, 248). A survey and evaluation of the relevant literature, including a critical evaluation of other proposed explanations for the employer size-wage effect is provided by Schwimmer (2007). For the present purposes it suffices however to observe that hiring standards seem to vary according to firm size, for whatever reason.
from a production line to individual assembly, would not affect monopsony wages but would increase selection wages, and the effect would be more pronounced for the group that supplies more elastically, which is the males.

A further difference between the theories relates to market clearance: Selection wages, just like other efficiency wages, allow for permanently uncleared markets, whereas monopsony wages would, in the dynamic monopsony case, imply rather invariant frictional and search unemployment only. This seems to be an advantage of the selection wage view, as unemployment in various labor markets seems often to be trended. (The rate of unemployment in Germany rose from below 1% in 1965 to above 9% in 2005. This is hardly compatible with the view that all firms operate along their respective labor supply curves, as the monopsonistic theory would assume.)

Regarding empirical findings, the proposed theory implies that the hiring standard for the disadvantaged group is lower than that of the higher paid group. Some of the cited literature on monopsony discrimination estimates the monopsonistic effects with great ingenuity, yet this cannot be generalized to selection wages easily, as the selection effect will depend on the distribution of productivities of the workers, which is unknown. Further, short-run elasticities differ from long-run elasticities, especially if social multiplier effects and repercussions on social stereotypes are taken into account. Any empirical study of selection wages should take care of these complications, and incorporate the other discrimination mechanisms (taste discrimination and statistical discrimination) as well. A “calibration” of any model involving only a single mechanism appears problematic, to say the least. It seems more fruitful, as a first step, to compare jobs that are particularly vulnerable to differences in the workers’ productivities with jobs where differences between workers are less important – for instance repair workers vs. assembly

10 In the case of selection wages, this statement needs to be qualified, as the market is always cleared at the equilibrium hiring standard. “Excess supply” refers here a larger supply at a lower, but still feasible, hiring standard. In contrast to other efficiency wage theories, the selection wage theory would permit employment being unaffected by changes in supply or demand, see Schlicht (2005).

On the other hand, and contrary to a wide-spread view, efficiency wages do not imply unemployment. Job seekers in a particular labor market may not compete wages down for efficiency-wage reasons, yet may be employed in another market that is cleared at a lower wage rate, see (Topel, 1988, 525) and Schlicht (2005).
line workers. The selection wage theory would suggest that sex discrimination is more pronounced in the former. In a similar vein, occupational discrimination – differential pay for similar workers working in different occupations – could be expected.

7 Occupational and Regional Discrimination

The argument presented in this paper is not confined to sex discrimination. Rather, we would expect discrimination to occur in all segments of the labor market where supply behavior differs.

Such a generalization may be illustrated by pay differentials that are unrelated to the male/female distinction. Assume that there are three types of workers and two jobs. Denote the types of worker by $a$, $b$, and $c$, and the jobs $B$ and $C$. Assume further that the workers of type $a$ are indifferent between both jobs (they prefer “any”), while the workers of type $b$ strongly prefer $B$, and the workers of type $c$ strongly prefer $C$.

Assume that the distribution of productivities is the same for all three types of workers in both occupations. Assume further that the social value of the output generated in jobs $B$ and $C$ is the same, and suppose that demand is such that some type $a$ workers must work in $B$, and some in $C$. Given this symmetric setting, wages in $B$ and in $C$ should be identical to achieve social efficiency. The type $b$ workers should go take jobs $B$, the type $c$ workers should take jobs $C$, and the type $a$ workers should fill the remaining vacancies in both types of jobs. The hiring standards in all jobs should be the same.

In presence of a selection problem, this efficient outcome may be destroyed if job preferences are correlated with supply behavior. If type $b$ workers are more strongly motivated by money and type $c$ workers are more strongly motivated by the desire to do something socially useful, wages will influence the supply behavior of $b$ workers more strongly than of $c$ workers, and wage differentials will emerge, and will be augmented by the social multiplier. Jobs $B$ will become more attractive and type $a$ workers will prefer them. Firms offering type $B$ jobs will be enabled to increase the hiring standard, while the $C$ firms will reduce theirs. Some type $b$ workers will fail to meet the increased hiring standard for $B$ jobs while some type
c workers will be able to meet the lowered hiring standard for C jobs and replace type a workers who have moved to B jobs.

Alluding to current stereotypes we may think of B as referring to banking, and C as pertaining to care, and we end up with an allocation that is socially inefficient. The bankers are overpaid relative to care workers, the work of the care workers underpaid. In so far as social esteem is connected to income and hiring standards, simply the fact that bankers, as a class, are more money-oriented and the care workers, as a class, are more task-oriented, will foster the bankers being socially esteemed and care workers looked down on. The example may serve to illustrate the thought that the selection wage mechanism may not only trigger sex discrimination in the labor market, but may induce inefficiencies and socially problematic outcomes in many other directions.

Incidentally, the example helps to correct some important and wide-spread theoretical misconceptions. In the example, groups of workers with identical productivity distributions sort across occupations according to “productivity,” in the sense that the hiring standard in banking is more demanding than that in care. Yet wages are efficiency wages, and the sorting cannot be taken as an argument that labor markets works competitively and efficiently, or that wage differentials are compensating differentials, as Murphy and Topel (1987, 137; 1990, 231) wrongly assert. Further, all markets are cleared in the sense that supply equals demand at the prevailing hiring standards. Yet this is not an indication that “efficiency wage contracts are dominated by other market clearing arrangements that eliminate worker rents” (Murphy and Topel, 1990, 237). The bankers earn more than the care workers, and this is socially inefficient, in spite of cleared markets.

A similar argument may be advanced for regional wage discrimination. Let B stand for jobs in the metropolitan area and C for jobs on the countryside. Denote by b workers who live in the metropolitan area and its suburbs, by c workers living on the countryside, and by a workers who live on the outskirts of the metropolitan area, and the earlier argument applies because firms in the metropolitan area will compete more fiercely for qualified workers, as workers will move more readily

---

11 Emmi Schlicht has suggested that the argument can be extended easily to non-monetary job features. If the attitudes of type c workers center on the social importance of the work they do, they will be less influenced by the general working conditions, and the firms will be less induced to offer nice workplaces and good equipment.
between them than between firms on the countryside, where mobility is reduced by larger distances between workplaces.

Fighting occupational wage discrimination of the type discussed above appears quite difficult. If attitudes towards work are correlated with attitudes towards money, and if we allow for a free choice of occupation, the analogues to equal pay enforcement or to the Scandinavian solution of assimilating social roles would be infeasible, since it is hard to imagine that we can enforce the same attitudes towards money across all social positions. It is therefore necessary to rely on other instruments, like collective bargaining or redistributive taxation, to bring wage differentials closer to compensating differentials, thereby enhancing both efficiency and fairness.

Conclusion

The selection wage theory of discrimination ties discrimination to different supply behavior of the groups involved. Firms set wages in order to attract more qualified applicants. As women respond less to wage differentials than men, firms can use the wage mechanism more effectively for men than for women. This brings about a higher market wage level for men than for women.

Empirical support for the view emerges from combining empirical studies of monopsonistic discrimination with empirical findings about the variability of hiring standards.

The argument is not confined to issues of sex discrimination; rather, it is of relevance for all labor markets where labor heterogeneity is important and differences in supply elasticities play a role. If this is the case and supply elasticities vary systematically across occupations, this may induce wage differentials between occupations that are both inefficient and unfair.

Appendix 1. Elasticities of Market Supply and the Firm’s Supply

The labor market literature considers as a stylized fact that labor supply elasticity of males is very small while the labor supply elasticity of women is considerable.
Alesina et al. (2006, 22 and Table 1.7), in their survey of various empirical studies, conclude that male supply elasticities range between $-0.07$ and $-0.35$, while the median estimate for women’s supply elasticity seems to be about 1. These observations seem to invalidate the assumption than men react more strongly to wage differentials than women, yet this is not the case: the studies relate to market supply elasticities, whereas the argument of this paper (along with the cited studies on monopsonistic wage differentials) refer to the supply elasticities, as seen by the firm. As noted by others, market level elasticities and firm-level elasticities may differ (Hirsch et al., 2008). The following note is aimed at clarifying the theoretical point.

Consider a typical firm operating in a market. Given its hiring standard, its labor supply is a function of its wage offer $w$ and the wage offers made by other firms operating in the same labor market. Summarize these other offers by the market wage rate $W$. We can characterize the labor supply to the firm as

$$x = \phi(w, W).$$  (19)

A natural assumption would be that this labor supply is increasing in the firm’s wage offer $w$ and decreasing in the market wage level that captures outside opportunities:

$$\frac{\partial \phi}{\partial w} > 0, \quad \frac{\partial \phi}{\partial W} < 0.$$

Denote the appropriately signed corresponding elasticities by

$$\eta = \left. \frac{\partial \phi}{\partial w} \right|_w x > 0$$

and

$$\chi = -\left. \frac{\partial \phi}{\partial W} \right|_w x > 0.$$

We refer to $\eta$ as the firm’s supply elasticity and to $\chi$ as the cross elasticity.

The market supply $X$ is obtained if the wage offers of all firms move in tandem with the market wage rate. We write

$$X = \phi(W) = a \phi(bW, W)$$  (20)
where \( a \) and \( b \) denote factors of proportionality: market supply is \( a \) times the typical firms’ supply when the typical firm sets its wage \( w \) proportionate to the market wage \( (w = bW) \).

The market supply elasticity is

\[
\epsilon = \frac{\partial \phi}{\partial W} \frac{W}{X}.
\]

From (20) we obtain \( \epsilon = \eta - \chi \). Re-arranging terms yields

\[
\eta = \epsilon + \chi.
\]  

(21)

Consider now the case of two types of workers, men and women, and super-script them by \( m \) and \( f \). From (21) it can be seen that we can have a case where the firm’s supply elasticity for women is less that that for males \( (\eta^f < \eta^m) \) while women’s market supply elasticity is larger \( (\epsilon^f > \epsilon^m) \). The requirement is that the difference between the market supply elasticities is more than offset by the cross elasticities: \( \epsilon^f - \epsilon^m < \chi^m - \chi^f \). A possibility would be \( \epsilon^f = 1, \epsilon^m = 0, \chi^f = 2, \chi^m = 4 \) amounting to \( \eta^f = 3 \) and \( \eta^m = 4 \). In the paper this is expressed by the phrasing that males react more strongly to wage differentials than females.

Appendix 2. Parameters Used For The Illustrations

All graphs use the function

\[
f(\theta, w, W) = \left(1 + \log\left(\frac{w}{W}\right)\right) \beta(\theta - 0.5)
\]

over the range \( \theta \in [0.5, 1.5] \), where \( \beta() \) denotes the probability density function of the Beta distribution with parameters \( (3, 5) \).

Figure 1 (a) uses \( \frac{w}{W} = 1 \), Figure 1 (b) uses \( \frac{w}{W} = 1.2 \).

Figure 2 (a) uses \( \frac{w}{W} = 1 \) and \( s = 0.7 \). This implies \( n = 0.85 \) and \( a = 0.92 \).

Figure 2 (b) in addition uses \( \frac{w}{W} = 1.4 \) which implies and \( s = 0.8 \) and \( a = 0.97 \).

Figure 3 (a) gives the productivity curve around \( W = 1 \). The competitive wage is \( w_0 = 0.86 \). Figure 3 (b) adds the case \( W' = 1.2 \) with \( w'_0 = 0.91 \). Figure 3 (c) gives the efficiency wage \( w^* = 1.1 \) resulting at \( p = 6.2 \).
References


Linked Employer-Employee Data from Germany. Working Papers 1111, Princeton University, Department of Economics, Industrial Relations Section. URL http://ideas.repec.org/p/pri/indrel/1111.html.


