NBER WORKING PAPER SERIES

A COMPETITIVE THEORY OF MONOPOLY UNIONISM

Edward P. Lazear

Working Paper No. 672

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

May 1981

The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #672
May 1981

A Competitive Theory of Monopoly Unionism

ABSTRACT

This paper sets up a microeconomic theory of labor unions. It discusses their formation and goals, their hierarchical structure, and the nature of rent distribution. The theory provides predictions for the probability that an industry or occupation will be unionized, the proportion of that industry that will be unionized, and observed wage differentials within that industry. It discusses the way that those values change in response to changes in the supply of labor, demand for labor, cost of organizing the union, and cost of defeating the union. Institutions such as featherbedding, fringe benefits, and seniority are rationalized in this framework. The model is consistent with competitive factor and product markets.

Edward P. Lazear
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637

(312) 753-4525

The last two decades have witnessed the growth of an increasingly rigorous brand of labor economics. The theories of labor supply, labor quality, information in the labor market, and labor contracts have received a significant amount of attention and the construction of theories with microeconomic foundations has led to a greater understanding of the phenomena. One area for which this has not occurred is the analysis of labor unions and how they function within the labor market. Models of union behavior treat the union rather than the individual in it as the basic unit of observation and suggest some behavior for that entity. This paper will adopt a micro-level approach and I will argue that one can go surprisingly far toward the analysis of union behavior by employing a simple model where workers and firms are the rational basic unit.

The goal is to predict which industries, occupations and time periods are most likely to be characterized by strong unions and to analyze the behavior of unions under a variety of circumstances. A central theme is that it is unproductive to think of the union as having an objective function.The point is perhaps best understood by considering some puzzles which arise when the union, rather than its more fundamental elements, is the basic unit of analysis. First, most models of union behavior allow a labor market which does not clear. Some workers who would like to work are turned away from union jobs. What happens to those workers? How do they influence the union's strategy? How can a labor market which does not clear be analyzed? Second, if not all firms in an industry are unionized, but union and nonunion firms sell the same product in an often competitive product market, then firms face different labor costs. How can they coexist selling the product at the same price? Third, given the highly hierarchical nature of most unions where seniority plays a major role, why don't the least priviledged workers within the union join forces with nonunion workers to undermine the power of the union? In particular, the young workers could form their own union, making employers and themselves apparently better off while destroying any union equilibrium. Why doesn't this happen more often? Finally, few models (Rosen [1969] is an exception) allow the firm to play an active role. In most, the labor demand curve is taken as given, and the union selects some wage which is consistent with maximizing its objective function. There is no allowance by the union, or the analyses in general, for the possibility that the firm will react differently to different wage demands except as expressed by labor demand.

This paper starts from utility and profit maximization by worker and firm respectively and builds a model which deals with these issues explicitly. Labor markets clear in the sense that anyone can enter the industry and claim a nonunion job, but there are queues for union jobs. The length of the queue is the direct result of union and firm behavior and the union takes this into account when announcing its wage demand. The product market clears as the firm pushes its anti-union activity to the point where costs are equalized across union and nonunion firms. The life-cycle nature of union benefits provides the young workers with an incentive to remain "true" to the union at the expense of short run gain. Since the young inherit the seniority rights from retiring workers, all workers are better off by this deferred payment structure. Finally, the firm plays a central role in determining the outcome of the union's efforts and its behavior is taken into account by the optimizing workers who potentially comprise the union.

"Union" is defined here as a collection of workers who act together to call out wage (and later, quantity) demands to firms. Firms buy labor from the union at the union wage or, at some cost, fight the union and buy labor in the competitive "fringe" (which may be much larger than the union sector). Product markets are competitive, and the equilibrium that emerges is a well-defined blend of competitive and monopoly, not unlike the dominant-firm-with-competitive fringe equilibrium found in the industrial organization literature. This simple structure yields a stable equilibrium which varies in unambiguous directions as the relevant observable parameters are altered.

The major implications of the model are:

1.) Union and nonunion firms earn the same profits so product market equilibrium is viable.

2.) The more elastic is the supply of labor to an industry or occupation, the lower is the probability that a union will exist in that industry or occupation.

3.) As the demand for labor becomes more convex, the probability that some proportion of the labor force will be unionized rises. Contrary to Marshall's assumption, inelasticity of demand for labor does not imply an increase in union power manifested by an increase in the union membership or wage differential.

4.) Inelasticity of product demand increases the likelihood that the industry will be unionized.

5.) As the cost of running a union and enforcing wage demands rises, the probability that a union exists falls. But given that one does exist, the union wage and wage differential increases and the proportion in the union tends to decrease as costs rise. A corollary is that anti-union legislation results in a smaller likelihood of a union, and a smaller proportion of union workers within a unionized industry, but larger union wage and wage differential. Also, as a corollary, as the attrition rate rises, the occupation is less likely to be unionized, but will have a higher wage differential if unionized.

6.) Wage demands are set by a centralized "national" union rather than by the "local" on a firm-by-firm basis. In addition, nationals may play an important role in inter-local union transfers through strike and pension funds.

7.) Union certification becomes less likely as the definition of the relevant electorate broadens. Further, a broader definition makes workers, as well as firms (although not necessarily union leaders), better off.

8.) Under reasonable circumstances, the observed wage differential overstates rather than understates the true effect of a union on wage rates and the overstatement is smallest for industries or occupations where the proportion of union workers is close to zero or one. In addition, there is no straightforward connection between wage differentials, proportion in the union, and union power.

9.) "Featherbedding" is both rent maximizing and Pareto Optimal. Further, a union which can select quantity as well as price selects the pre-union employment level.

10.) Union workers within an industry or occupation will tend to be older than the nonunion workers. Further, as the proportion in the union increases, the average age of union workers is nonincreasing, the average age of nonunion workers is nonincreasing, and the difference between the age of union and nonunion workers first increases then decreases with the proportion of unionized workers.

11.) Union workers take a greater part of their compensation in the form of fringes than do nonunion workers.

12.) Although older workers receive greater benefits from the union than younger workers, young workers have no incentive to break the coalition. However, in the absence of any initial union set-up costs, once-and-for-all transfers to the older workers make them more likely to vote for a union than young workers.

13.) Union-nonunion wage differentials move countercyclically.

I. The Model

Consider an industry comprised of S competitive firms each of which has a demand for labor given by $L = d(W)$ where W is the wage rate and L is the number of workers in the firm. There are $R(W)$ workers in the occupation or industry. Assume (relaxed below) that $R(W) = R$ so that labor is supplied inelastically to the occupation.[1] Also assume that the demand for output is perfectly elastic so that second-order product market effects can be ignored. The structure is to consider first a one period setting, where results are easier to obtain and then to prove that these results hold in a multiperiod context with some modifications.

The Opportunity Locus:

Risk neutral workers can band together to form a "union" which calls out a union wage, W_U. An individual firm faced with the union demand can either pay W_U and then choose to hire $d(W_U)$ workers or at some fixed cost, C_i, which varies across firms, can defeat the union and pay wage W_N, hiring $d(W_N)$ workers. C_i can be thought of as the cost of employing enough "union busters" to defeat the union or as contributions to an employee benefits fund which appease the current work force. Alternatively, if unions have beneficial effects on productivity as Freeman [1976] and Brown and Medoff [1978] have argued, C_i is the cost of foregone productivity effects of the union. Let $C_i \sim g(C_i)$ with distribution function $G(C_i)$.[2] That workers can impose a cost C_i, on a firm endows them with a property right. The National Labor Relations Act may increase the cost associated with defeating the union, but even in the absence of this law, workers impose some costs on firms by disruptive actions and thereby retain some property right.

Let the firm have a standard concave production function. The firms profit function depends upon the price of output, price of capital and wage rate. Initially assume that demand for the product is perfectly elastic so that the first two prices are given and invariant across firms. Suppress the prices of capital and output $\Pi = \Pi(W)$. If the firm hires union labor, the firm's profits are $\Pi(W_U)$. If the firm hires nonunion labor, the firm's profits are $\Pi(W_N) - C_i$. The firm chooses to fight the union if

$$\Pi(W_U) < \Pi(W_N) - C_i$$

or if

$$\Pi^*(W_U, W_N) \equiv \Pi(W_N) - \Pi(W_U) > C_i .$$

For any given W_U, W_N combination then $G(\Pi^*(W_U, W_N))$ of the firms will find it more profitable to be nonunion firms.[3] Therefore the demand for labor by nonunion firms is $S[G(\Pi^*[W_U, W_N])]d(W_N)$ and the demand for labor by union firms is $S[1-G(\Pi^*[W_U,W_N])]d(W_U)$. The market equilibrium condition is that demand for labor equal the supply of labor:

$S[1-G(\Pi^*[W_U, W_N])]d(W_U) + S[G(\Pi^*[W_U, W_N])]d(W_N) = R$ or in per-firm notation,

$$(1) \quad [1-G(\Pi^*[W_U, W_N])]d(W_U) + G(\Pi^*[W_U, W_N])d(W_N) = R/S .$$

In this one period setting, the R workers are locked into the industry so eq. (1) defines the union's "opportunity locus." For any W_U that the union chooses, a W_N will result which selects some proportion of the firms as union firms consistent with the condition that supply equals demand. A higher W_U affects W_N in two ways: First, it increases the proportion of firms that fight the union. That there are fewer union firms implies a greater supply of labor to the nonunion sector, but also that there are a larger number of nonunion firms which implies a greater demand for labor in that sector. Second, higher W_U decreases the number of workers hired by each of the union firms, sending the spillover into the nonunion sector. As long as demand curves are downward sloping, the basic shape of the opportunity locus will <u>always</u> be as shown in figure 1. In particular, it has its starting point at $[W_C, W_C]$ where W_C, the competitive wage, is given by $d(W_C) = R/S$. It is always negatively sloped initially, positively sloped as W_U gets large, below W_C so that the nonunion wage is inferior to the pre-existing competitive equilibrium, and it asymptotes to W_C.[4]

Consider the last statement first. As W_U goes to infinity, $G(\Pi^*(W_U, W_N))$ goes to 1, i.e., all firms choose to fight the union. Then (1) becomes

$$d(W_N) = R/S$$

so $W_N = W_C$. The intuition is clear. If the union chooses such a high wage that all firms fight the union, then all workers and all firms are in the nonunion sector so the situation there is identical to the initial competitive labor market.

It is equally clear that as W_U goes to W_C from above (no worker or union will choose $W_U < W_C$), W_N goes to W_C: From (1), we require

$$\lim_{W_U \to W_C} [1-G(\Pi^*[W_U,W_N])]d(W_U) + G(\Pi^*[W_U,W_N])d(W_N) = R/S$$

$$\text{or } [1-G(\Pi^*[W_C,W_N])]d(W_C) + G(\Pi^*[W_C,W_N])d(W_N) = R/S$$

But this is just a convex combination of $d(W_C)$ and $d(W_N)$. Since $d(W_C) = R/S$, only $W_N = W_C$ will avoid a contradiction.

The intuition here is also clear. As W_U shrinks to W_C, union firms are demanding no less than their competitive share of labor and extracting workers from the nonunion sector in proportion to their numbers. Therefore, the number of workers left in the nonunion sector does not exceed the amount that these firms employ in competition and so the wage rate there must be the competitive one.

Since $W_U > W_C$ implies that union firms demand less than their proportionate share, the spillover of workers to the nonunion sector must exceed the number that those firms employ in competition. This means that $W_N < W_C$ throughout. Given that $W_N = W_C$ as W_U approaches W_C or infinity, it must be the case that $\left.\frac{dW_N}{dW_U}\right|_{OL}$ (OL opportunity locus) starts out negative and ends up positive. This can also be seen by differentiating (1) totally to obtain $\left.\frac{dW_N}{dW_U}\right|_{OL}$:

$$(2) \qquad \left.\frac{dW_N}{dW_U}\right|_{OL} = -\left\{\frac{d'(W_U)[1-G(\Pi^*)]+g(\Pi^*)(\Pi^*_1)[d(W_N)-d(W_U)]}{d'(W_N)G(\Pi^*)+g(\Pi^*)\Pi^*_2[d(W_N)-d(W_U)]}\right\}$$

Recalling that $\Pi^*(W_U,W_N) = \Pi(W_N) - \Pi(W_U)$ and noting that the derivative of a profit function with respect to the price of an input is the negative of the demand for that input,

$$\Pi_1^* = -(-d(W_U)) = d(W_U) > 0$$

and

$$\Pi_2^* = -d(W_N) < 0$$

so we can rewrite (2) as

$$(3) \qquad \left.\frac{dW_N}{dW_U}\right|_{OL} = -\left\{\frac{d'(W_U)[1-G(\Pi^*)]+g(\Pi^*)d(W_U)[d(W_N)-d(W_U)]}{d'(W_N)G(\Pi^*)-g(\Pi^*)d(W_N)[d(W_N)-d(W_U)]}\right\}.$$

The denominator is negative since $d' < 0$, $g > 0$, $G > 0$, and the sign of the numerator changes. As $W_U \to W_C$, $G(\Pi^*) = 0$ and $d(W_N) - d(W_U) = 0$ so $dW_N/dW_U \big|_{OL} = -\infty$. As $W_U \to \infty$, $G(\Pi^*) = 1$ if we define $d(\infty) = 0$, then $dW_N/dW_U \big|_{OL} = 0$. The reason that W_N rises as W_U approaches infinity is that although more workers are being thrown into the nonunion sector, firms are switching from union to nonunion at an even more rapid rate, bidding up the price of labor there.

An Alternative Derivation of the Opportunity Locus and Product Market Equilibrium:

The reader may find artificial the assumption that firms are identical in all respects except for the cost of defeating the union. But the "union busting" function is easily separated from the firm and the analysis is unchanged. Let firms be identical in all respects. Let there be a group of S potential union busters whose alternative use of time is $C_i \sim g(C_i)$ across the S union busters. At any W_U, W_N pair, firms will pay up to $\Pi(W_N) - \Pi(W_U)$ to employ a union buster so the demand for union busters is perfectly elastic at price $\Pi(W_N) - \Pi(W_U)$. The number of union busters who supply their services is then

$SG(\Pi(W_N) - \Pi(W_U))$ or $SG(\Pi^*(W_U, W_N))$ so $SG(\Pi^*)$ firms are nonunion and

$S(1-G(\Pi^*))$ are union and we are back to eq. (1). All firms are identical and all earn the same profit rate equal to $\Pi(W_U)$ since nonunion firms receive $\Pi(W_N)$ but pay out $\Pi(W_N) - \Pi(W_U)$ to the union buster.

Union and nonunion firms exist in the same industry and there is no tendency for the high wage union firms to be driven out of business. Even if we think of the C_i as reflecting different costs of beating the union across firms this proposition holds. The reason is that the enterpreneur or other scarce factor that is responsible for that firm's being an effective union buster captures the rent, $(\Pi(W_N) - C_i) - \Pi(W_U)$ because union firms are willing to pay up to this amount for the scarce factor's services. So $(\Pi(W_N) - C_i - \Pi(W_U))$ goes to the scarce factor, C_i is the direct cost of beating the union so the nonunions firms profit is again $\Pi(W_U)$ and product market equilibrium is maintained. Union and nonunion firms exist in the same product market, each earning the same level of profit.

The Indifference Curves:

A potential union consisting of all R workers takes the opportunity locus as given. It can select any W_U, but this will imply a particular W_N. In addition, the choice of W_U affects the proportion of workers who will be employed in the union sector for two reasons. First, it alters the number of firms who choose to fight the union. Second, it changes the number of workers employed by each union firm. Since, in this section, all workers are identical ex ante, the probability that a given worker will be employed by a union firm and receive wage W_U is

(4) $$P \equiv \frac{S(1-G[\Pi^*(W_U, W_N)])d(W_U)}{R}$$

(the number of union workers divided by the total labor supply). Every $[W_U, W_N]$ pair implies a P. However, only $[W_U, W_N]$ pairs which lie on the opportunity locus correspond to feasible P, W_U, W_N combinations. (The probability that a union job is obtained varies by worker characteristic, of course, and this is the subject of the last half of this paper.)

Workers are identical and risk-neutral ex ante so each has the same objective function: maximize expected wealth. If it costs Z per union worker to administer the union, i.e., to hire the union leader, strike or enforce demands through other methods, then the workers objective function is to chose W_U so as to solve

$$(5) \qquad \max_{W_U} \; P(W_U - Z) + (1 - P)W_N$$

subject to the constraint implied by (1), i.e., subject to being on the opportunity locus. Note that this answers the question as to which members does the union represent. Ex ante, all R workers are represented. Ex post, PR workers are union members.[5]

This is a straightforward maximizaiton problem, but insight can be gained by considering it in two stages. We construct the union's (i.e., each worker's) indifference curves from (5) and select W_U such that the indifference curve is tangent to the opportunity locus.

The indifference curve corresponding to any utility (wealth) level K is

$$(6) \qquad K = P(W_U - Z) = (1 - P)W_N$$

or

$$(7) \qquad W_N = \frac{K - P(W_U - Z)}{1 - P}$$

Indifference curves are shown in figure 1. They almost always display this shape. In particular, they start at the 45° line at $K + Z$, have negative slope initially, then positive slope, generally have no inflexion points in the negatively sloped region, one in the positively sloped region and asymptote to K . Consider each in turn.

First, if $W_U = W_N$ then (6) says $K = P(W_N - Z) + (1 - P)W_N$ or $W_N = K + Z$. Second, as W_U goes to infinity, from (4), $P = 0$; all firms fight the union and union firms demand no workers. So from (7), $\lim_{W_U \to \infty} W_N = K$.

The slope of the indifference curve is obtained by differentiating (6):

$$(8) \qquad \left.\frac{dW_N}{dW_U}\right|_{IC} = - \left\{ \frac{P + \partial P/\partial W_U (W_U - Z - W_N)}{1-P + \partial P/\partial W_N (W_U - Z - W_N)} \right\} .$$

where $\partial P/\partial W_U$, $\partial P/\partial W_N$ obtained from (4) are

$$(9) \qquad \partial P/\partial W_U = \frac{S}{R} (1-G(\Pi^*))d'(W_U) - \frac{S}{R} g(\Pi^*)\Pi_1^* d(W_U)$$
$$= \frac{S}{R} (1-G(\Pi^*))d'(W_U) - \frac{S}{R} g(\Pi^*)[d(W_U)]^2 < 0$$

and

$$(10) \qquad \partial P/\partial W_N = \frac{S}{R} g(\Pi^*)\Pi_2^* d(W_U)$$
$$= \frac{S}{R} g(\Pi^*)d(W_N)d(W_U) > 0 .$$

The slope on the 45° line is given by

$$(11) \qquad - \frac{\frac{S}{R} d(W_U) - Z(\frac{S}{R} d'(W_U) - \frac{S}{R} g(0)d(W_U)^2)}{1 - \frac{S}{R} d(W_U) [1+Zg(0)d(W_U)]}$$

The numerator is positive. If Z, $g(0)$ or $d(W_U)$ is small then the denominator is positive so the slope of the indifference curve is negative.[6]

At the other extreme, as W_U approaches infinity, W_N is K , $P = 0$, $d(W_U) = 0$ so $\lim_{W_U \to \infty} \left. \frac{dW_U}{dW_N} \right|_{IC} = 0$.

The sense of the U-shaped indifference curve is this: Workers would like both higher W_U and higher W_N if nothing else were involved. But a higher W_U implies lower P . Initially, the value of an increase in W_U outweighs the loss associated with a reduction in P so indifference curves are negatively sloped. As W_U gets large and P gets small, however, the benefits to increased W_U are swamped by the loss resulting from a decreased P and workers view additional W_U as a bad, yielding positively sloped indifference curves.

Equilibrium:

Define the "critical indifference curve" as the one that yields the same utility as the competitive equilibrium. Using (7), the critical indifference curve is the one that has $K = W_C$ or

$$(12) \qquad W_N = \frac{W_C - P(W_U - Z)}{1-P} .$$

A union equilibrium exists with certainty if the critical indifference curve crosses the opportunity locus, since this implies that there exists some feasible W_U, W_N, P combination which yields an expected utility level higher than the one offered by competition. In figure 2, if IC_O were the critical indifference curve, then a union equilibrium would exist. It need not exist, however. If Z were very large, for example, so that it is expensive to run the union, an equilibrium is unlikely to exist. Although the critical indifference curve and opportunity locus both asymptote to W_C , the indifference curve may well lie everywhere above the opportunity locus. In this case, workers are better off accepting the competitive wage and not forming a union.

If a union equilibrium exists, i.e., if the critical indifference curve intersects the opportunity locus, the selection of an optimum union wage is given by the first order conditions of (5) that

$$(13)\qquad \text{a.}\quad \left.\frac{dW_N}{dW_U}\right|_{IC} = \left.\frac{dW_N}{dW_U}\right|_{OL}$$

$$\text{b.}\quad [1-G(\Pi^*)]\,d(W_U)+G(\Pi^*)\,d(W_N) = R/S\ .$$

Therefore, given a $G(\Pi^*)$ distribution, a $d(W)$ function and Z, the union optimum can be obtained. The solution described by (13a,b) is almost always an interior one, given the shapes of the indifference curves and opportunity locus. This implies that an occupation or industry will almost never be entirely unionized, if it is unionized at all. It always pays to leave some of those firms best able to "beat the union" out of the union sector, rather than choosing a union wage so close to W_C as to make it unprofitable for them to oppose. This seems to fit the stylized fact that unions rarely, if ever, organize all firms in an industry.

Finally, in this one-period context, no difficulty arises with respect to ex ante v. ex post preferences. Although union workers might wish to behave differently once they realize that they have won the lottery, this is not permitted by the one-period framework. Below, in the multiperiod context, discrepancies between ex ante and ex post winners and losers are resolved by assigning workers to the union on an age or seniority basis. The union jobs will always go to the older workers first so that over one's lifetime, each worker expects to be a loser at some time and a winner at others.

A key point is that even if ex post losers would like to bid with ex post winners for the union jobs, they are precluded from doing so. The firm can entertain these offers as a way to beat the union. As such, it carries with it cost C_i imposed by union workers, and therefore is already implicitly accounted for in $G(C_i)$, the distribution of costs of defeating the union. This model does not permit firms to costlessly accept the labor offer of a scab. Even in the absence of NLRB rules, it seems reasonable to view the hiring of scabs as carrying some real costs, perhaps larger to some firms than others.

A Note on Labor Market Equilibrium:

The labor market clears in the sense that aggregate supply of labor equals aggregate demand. Workers prefer the union jobs and in a multiperiod context (below) queue for them. But this market permits workers to enter the occupation in accordance with their own labor supply optimization and no artificial supply restrictions are required. The wage in the nonunion sector adjusts to clear the market which allows free entry.

Incidentally, herein lies the difference between this model and the dominant firm construct found in the industrial organization literature (see Cohen and Cyert (1975) and Carlton (1979), for examples). First, dominant firms, by restricting their supply insure that their own profits are maximized and second, commodities trade at the same price. In this model, the union does not have any power over the labor supply function and the wage in the nonunion sector lie below the union wage. Firms, unlike customers, are not permitted costlessly to buy from the lowest price sellers of labor. In the dominant firm model, C_i is zero for all consumers, and their ability to buy from the lowest priced seller insures that the price of the commodity is the same across sellers.

II. Extensions and Implications of the Model

The Supply of Labor:

For expositional convenience, we assumed that the supply of labor to the occupation or industry was perfectly inelastic. In this section we relax that assumption, allowing the labor force to be responsive to the actions of the union. An individual considers the expected utility from work in an occupation and compares it to the alternatives. By definition, if a union organizes an occupation, it does so because the expected utility of each member rises. Therefore let R, the number of workers in the occupation be given by $R(K)$ where K, defined in (6), is the expected utility of entering the occupation and $R'(K) > 0$. As the expected wage in the occupation rises, those whose comparative advantage previously lay elsewhere, are now induced to acquire the requisite skills for this occupation.

The main result is that a union is less likely to exist as the supply of labor becomes more elastic. Mechanically, this is because the opportunity locus shifts downward and the critical indifference curve shifts upward making an intersection of these curves less likely. The proof of these propositions is contained in appendix A. The intuition behind it is this: When the supply of labor is elastic, choosing a higher-than-competitive union wage induces more workers into the industry. As the result, the wage in the nonunion sector must fall more in order for nonunion firms to take up this larger residual labor force. This forces the opportunity locus downward. Similarly, since the probability that a given worker will obtain a union job is lower the more workers there are in the industry, a worker requires a higher nonunion wage for a given union wage to obtain the same level of expected utility. Thus, the critical indifference curve or the minimum combination of (W_U, W_N) which the worker views as preferable to the competitive wage, W_C, rises, making a union equilibrium less likely.

At the extreme, the supply of labor to the occupation could be perfectly elastic. This requires a particular kind of homogeneity in ability so that no workers are relatively better at some occupations than others. Under these circumstances, no union equilibrium exists because raising the expected wage above W_C brings about an infinite sized labor force so that the probability of obtaining a union job falls to zero and all workers prefer the competitive wage. To the extent that the union can restrict entry into the occupation, it gives some inelasticity to the labor supply curve and we are back to the case just analyzed.

This may be the difference between unions organized along craft (occupational) lines and those organized along industrial lines. To the extent that the long run supply of labor to an occupation is more inelastic than to an industry, unions are more likely to be successful in crafts. As long as it is easier for an individual to switch from industry i to j than it is for him to switch from occupation k to h, the proposition holds. Heterogeneity in talents across occupations makes this feasible, even for long run labor supply. Historically, craft unions were organized before industrial unions and there is some evidence (see Lewis (1963), Rees (1962)) that the former have been more successful than the latter.

The Demand for Labor:

Let us begin this section on a negative note: Although there are a number of statements that can be made about how the demand for labor affects union behavior, excluded from that set is the traditional (á la Marshall, Pigou, Hicks) and intuitive claim that unions have more power and are more likely to be successful the more inelastic is the demand for labor. The reason is quite simple. Although it is true that a given wage increase displaces a smaller number of union workers the more inelastic is the demand for labor, a given number of displaced workers drives down the nonunion wage by more, the more inelastic is the demand for labor. These spillovers must be absorbed by a nonunion sector firm with that same inelastic labor demand, implying that the wage rate in nonunion jobs must fall by more. This reduces the expected wage at the same time that the inelasticity increases it via higher union wages. The net effect is ambiguous, but in some cases, the effects are exactly offsetting so that inelasticity of labor demand does not affect the equilibrum.

Surprisingly, what is crucial is convexity of the demand curve. As the demand for labor becomes more convex, a given increase in the union wage results in a smaller displacement of workers. Also, as the demand for labor becomes more convex, a given displacement of workers into the nonunion sector reduces wages by a smaller amount there. As the result, the opportunity locus shifts upward, i.e., for a given union wage, a higher nonunion wage is available. Similarly, since the probability of obtaining a union job rises with the convexity of the labor demand curve, workers are as well off as before with a lower nonunion wage. Thus the critical indifference curve shifts down, i.e., at a given union wage, a lower nonunion wage is acceptable. Both of these forces increase the probability of obtaining a union equilibrium.[7] The proof of this proposition is contained in the last section of appendix 2.

An increase in the demand for labor can have almost any effect depending upon the shape of the new demand curve relative to the old. However, in the simple case where demand curves are linear and the increase takes the form of a parallel shift, there will be no effect on the probability of obtaining a union equilibrium nor on the union-nonunion wage differential. The formal proof is contained in the first part of appendix 2. This suggests that business cycle variations in union behavior and results are not caused by difference in the demand for labor per se. Findings such as those by Freeman (1980), Medoff (1979), and Blau and Kahn (1980) that union employment is more procyclic than nonunion employment must rely on inter-industry or interoccupational differences in product cyclical sensitivity for their explanation.[8]

Demand for Product:

These results suggests that the analyses of Marshall, Hicks, and Pigou are based on an inappropriate assumption. They argued that the more inelastic is the demand for labor, the more likely it is that unions raise wage. They proceed to discuss the conditions under which the demand for labor will be more inelastic. In the last section, we showed that the premise on which their analysis is based may be false. Yet, one of Marshall's implications, that union power increases with the inelasticity of product demand, holds. This is not because of the relationship between inelasticity of product and labor demands, but because of the relationship between inelasticity of product demand and convexity of labor demand.

The more inelastic is product demand, the more convex is the relevant labor demand curve. To see this, consider a firm which produces with labor plus one manager. In competitive equilibrium, the output price, P_{Q_0}, is defined by the minimum of the presumably U-shaped average cost curve which depends, in turn, upon the wage rate. If $D(P_Q)$ is the market demand for product, output $Q = f(L)$ given manager present, then equilibrium is the vector $[L, W, P_Q, S]$ which solves the following system:

$(S)L = R$ (Demand for labor equals supply of labor)

$Sf(L) = D(P_Q)$ (Supply of product equals demand for product)

$W = P_Q f'(L)$ (Firm hires labor to set VMP equal to wage)

$P_Q = \underset{Q}{MIN}[\text{Average Cost } (W, \text{fixed cost}, f(L))]$

(firms earn zero profits).

Starting at equilibrium shown in figure 2 by $[W_C, R/S]$, let the union raise the wage rate to W_U. If all else were the same, the firm reduces labor demand to l_0 . But the increase in the wage to W_U shifts the average cost curve upward. If all firms were unionized, this would raise the equilibrium price of output, P_Q . But even if all firms are not unionized, this must happen. Since this move raises the value of a "union buster," the salary of the manager in the nonunion firm is bid up to the point where the minimum of the average cost in the nonunion firm equals that in the union firm. (Because this is a fixed cost to the nonunion firm, output there will rise while the higher marginal cost induces output to fall in the union firm.) When P_Q rises to its new level, the demand for labor shifts from $d(W)$ to $\tilde{d}(W)$ so each union firm employs l_1 rather than l_0 of labor. Similarly, nonunion firms employ l_2 of labor and W_N is the solution to equation (1). The relevant demand for labor curve is then AB for union firms and BC for nonunion firms which adds convexity to the demand for labor. Further, the more inelastic is the demand for output, the larger is the increase in price P_Q which makes the labor demand even more convex. But as argued above, increased convexity increases the likelihood of a union equilibrium so less elastic product demand results in a higher probability of a union.[9]

The Costs of Operating the Union:

It has already been noted that as the cost of running the union and enforcing wage demands, Z , rises, the likelihood of a union equilibrium diminishes. Here it is shown that if a union equilibrium does exist, the observed union wage and wage differential will be larger when operating costs are higher. In addition, the proportion of workers in union jobs tends to fall as operating costs rise.

Recall that the shape of the opportunity locus is independent of Z . Consider the $[W_{U_0}, W_{N_0}]$ optimum given some Z_0 . If $\left.\frac{dW_N}{dW_U}\right|_{IC}$ decreases with Z then the indifference curve for $Z_1 > Z_0$ through $[W_{U_0}, W_{N_0}]$ must cut the opportunity locus from above. This implies that $W_{U_1} > W_{U_0}$. To see this analytically, differentiate (8) with respect to Z :

$$(14)\quad \frac{d\left(\left.\frac{dW_N}{dW_U}\right|_{IC}\right)}{dZ} = \frac{[1-P + \partial P/\partial W_N(W_U - W_N - Z)]\partial P/\partial W_U + [P + \partial P/\partial W_U(W_U - W_N - Z)](-\partial P/\partial W_N)}{[1-P + \partial P/\partial W_N(W_U - W_N - Z)]^2}$$

$$= \frac{(1-P)\ \partial P/\partial W_U - P\ \partial P/\partial W_N}{[1-P + \partial P/\partial W_N(W_U - W_N - Z)]^2} < 0 .$$

So as Z rises, the slope of the indifference curve through any point falls. This implies the optimal W_U increases with Z .

Further, $W_U - W_N$ increases as well since the slope of the opportunity locus is less than one (see (2)). This implies that P will tend to decline as Z rises. There will be fewer union workers, but each union worker will earn a higher union wage to compensate for the higher costs of administering the union. (See appendix 3 for the formal discussion.)

One can also imagine varying the costs of opposing the union. For example, laws which make it more difficult to defeat the union would shift the $G(C_i)$ distribution rightward. Alternatively, if the shape of the $G(C)$ distribution does not change, pro-union legislation can be thought of as lowering Z, the cost of organization. Now for a lower level of expenditures by workers on union organization, the same distribution of costs of defeating the union prevails. The implication is straightforward: Pro-union legislation, by reducing Z for a given distribution of C_i, increases the probability of obtaining a union equilibrium, increases the proportion in the union, and lowers the union wage and wage differential. This is analogous to the situation in monopolistic product markets where cost saving technology which lowers the marginal cost function increases the firm's output and profits, but results in a decrease in price. It carries with it the somewhat paradoxical result that prounion legislation reduces the distortion in unionized industries.

To the extent that Z is lower in occupations or industries in which firm size is relatively large, unions are more likely to exist there. However, if a union equilibrium does exist in a small firm industry, other things equal, the wage differential will be larger and the proportion unionized smaller in this industry. The empirical implication is that the probability of there being any union workers in an industry and the proportion of the work force holding union jobs rises with firm size. However, among at least partially unionized industries, the union-nonunion wage differential varies inversely with average firm size.

Nationals and Locals:

The discussion has been in the context of a union selecting a wage for the entire industry or occupation. Thus, bargaining takes place at the "national" level where "national" is defined as the market over which workers are perfect substitutes. Now consider another possibility. Start with a competitive industry and allow each firm's R/S workers to vote separately on whether or not that firm should be unionized and then to select the union wage. The result is that locals do not maximize workers' expected wealth so that wage setting should occur at the national level. This implication is not new, but the usual reason for it relies on the relationship between bargaining power and union worker solidarity. This is not what is operating here. Because one local union's behavior affects market price and the aggregate probability that a worker obtains a union job in a way different from that perceived by the local, an inferior solution results. Further, it is novel that the local sometimes establishes a wage rate that is too low rather than too high.

This section also lays the groundwork for considering behavior by workers in a union firm versus that by those in nonunion firms. As such it is worthy of some detail.

Consider a wage W_N. The workers in an individual firm or local union, unlike the national, take this as given and assume that they cannot influence it. Also, the probability of being in a union that is relevant for workers in firm i is

$$(15)\qquad \tilde{P} = \begin{cases} 0 & \text{if } \Pi^*(W_U, W_C) > C_i \\ d(W_U)/d(W_N) & \text{if } \Pi^*(W_U, W_C) < C_i \,. \end{cases}$$

That is, if W_U is sufficiently high so that firm i opts to defeat the union (say, by bribing the vote counter), then the probability of obtaining the union wage is zero. If the firm accepts the union, then the work force of that firm will be reduced from $d(W_C)$ to $d(W_U)$ workers so $[d(W_U)]/[d(W_C)]$ will retain their jobs while the rest are forced into the nonunion sector.

The equation for the local union's indifference curve is then

$$(16) \qquad \tilde{W}_N = \frac{K - \tilde{P}(W_U - Z)}{1 - \tilde{P}} .$$

By differentiating (7) with respect to P, we obtain

$$(17) \qquad \frac{dW_N}{dP} = \frac{K + Z - W_U}{(1 - P)^2} < 0$$

since in the relevant region, $W_U > K + Z$.

Note also that

$$\begin{aligned} \tilde{P} \equiv \frac{[d(W_U)]}{[d(W_C)]} &\geq \frac{[d(W_U)]}{[d(W_C)]} [1 - G(\Pi^*)] \\ &\geq \frac{d(W_U)}{R/S} [1 - G(\Pi^*)] \\ &\geq \frac{S}{R} [1 - G(\Pi^*)] d(W_U) \equiv P . \end{aligned}$$

Since $\tilde{P} \geq P$, $\tilde{W}_N \leq W_N$ so the critical indifference curve for myopic locals lies below that for the national. Thus, a local union might form even if the competitive situation is better for all workers. Since all locals are ex ante alike, these unions will learn of their mistakes at best ex post. Whether they decertify the union, render it impotent or persist in long run myopic behavior is not clear and is a standard prisoner's dilemma problem. As such, the NLRB rules which allows the choice of union status by workers in elections generally held on a per firm basis, results in too much unionization in the sense that it does not maximize worker's wealth. Because the workers in each firm do not take their effects on the market into account, they are too inclined toward unions.

This suggests the implication that the wider the definition of the voting population, the more likely is the unit to recognize market effects and therefore the less likely is the establishment of a union. So an NLRB policy which broadened the population over which it held certification elections, would make workers, as well as firms (although not necessarily union leaders), better off because it alters the prisoner's dilemma nature of the payoff structure.

Even if a union equilibrium does yield a higher expected wealth level than the competitive labor market, it is obvious that the Nash equilibrium when locals choose W_U will deviate from the equilibrium obtained when a national sets W_U. Since each local takes W_N as given and cares about $\tilde{P} = \frac{d(\tilde{W}_U)}{d(\tilde{W}_C)}$ rather than $P = \frac{S}{R}[1 - G(\Pi^*)]d(W_U)$, the conditions for a Nash equilibrium when locals choose wages are

$$(18)\quad a.\quad \left.\frac{dW_N}{dW_U}\right|_{IC} \equiv \frac{\tilde{P} + \partial\tilde{P}/\partial W_U\,[W_U - Z - W_N]}{1 - \tilde{P} + \partial\tilde{P}/\partial W_N[W_U - Z - W_N]} = \left.\frac{dW_N}{dW_U}\right|_{OL}$$

or

$$\tilde{P} + \partial\tilde{P}/\partial W_U[W_U - Z - W_N] = 0$$

and

$$b.\quad [1 - G(\Pi^*[W_U, W_N])]d(W_U) + G(\Pi^*[W_U, W_N])d(W_N) = R/S\ .$$

(18a) says that since locals assume that they have no effect on the wage rate, the relevant opportunity locus, as they see it, is $W_N = W_N^*$ with $dW_N/dW_U\big|_{OL} = 0$. (18b) merely repeats the condition that the solution, to be rational, must lie on the opportunity locus. The conditions for the national's equilibrium were given in (13a) above. Therefore, the equilibrium will differ except in the rare case when the national's optimum is located at the minimum point on the opportunity locus.

It is interesting to note that the local's choice of W_U , and W_U in equilibrium, is not always above the national's W_U equilibrium. Consider figure 3. Suppose that the national's equilibrium were at A . Each local would try to go to B which is unobtainable of course, and the equilibrium would be at C where

$$\left.\frac{dW_U}{dW_N}\right|_{IC_{Local}} = 0 , \quad \text{and} \quad W_U, W_N$$

are on the opportunity locus. In this case W_U would be higher if locals set wages than if nationals set wages. Alternatively, if the national's equilibrium were at D , locals would like to go to E , and equilibrium would result at F with W_U lower than the national's choice of W_U . The reason for the difference is the following: Local unions do not take into account the effect that their actions have on the nonunion wage by driving workers and firms into the nonunion sector. At A , the effect of raising W_U is to lower the nonunion wage so that by ignoring it, locals set too high a wage. At D , the effect of raising W_U is to raise the nonunion wage so ignoring this causes locals to set too low a wage.

Locals, because they ignore spillovers, choose a wage rate which does not maximize the expected wealth of workers in the entire occupation. Although any one local may be better off as the result of all unions behaving myopically, the rest are sufficiently worse off to reduce average wealth. This suggests that a national union which sets the optimal industry wage could make all workers better off ex ante. Therefore wage setting should be done at the national level. (Recall that "national" is used loosely here. It refers to the relevant labor market defined by the pool over which workers are perfect substitutes.)

To the extent that some locals may lose relative to others by this arrangement, it may be necessary to couple this with transfers from one local to another. A centralized strike or pension fund, for example, which doles out benefits in accordance with some prearranged formula, might be such a transfer mechanism.

Threat Effect and the Measurement of Union Power:

Researchers have worried that the observed union differential may understate the true effect of the union on wage rates because nonunion firms, in attempting to discourage their workers from becoming unionized, pay more than the competitive wage.[10] As the result of the spillovers into the nonunion sector, however, it is known that the effect may well go on the other way. Novel is that the understatement or overstatement of the true differential bears a particular relationship to the proportion in the union.

Let us continue to think of C_i as being resources spent to hire a union buster or bribe a vote counter. Consider figure 1. At any given wage rate, W_U , the true amount by which unions raise wages is $W_U - W_C$, measured as the vertical distance between the 45° line and the horizontal line at W_C . The observed wage differential, however, is $W_U - W_N$ or the vertical distance between the 45° line and the opportunity locus. Since OL is everywhere below W_C , the observed differential overstates rather than understates the true effect of the union on wage rates. This is because the effect of a union is to depress the nonunion wage. Further, since the difference between W_C and W_N shrinks as W_U approaches either W_C or ∞ , the overstatement of the true effect is smallest when the proportion of the workers in the union is close to zero or to one.

Crucial here, of course, is the assumption that C_i does not become part of the observed wage. Thinking of C_i as salary to a union buster, a bribe for an official or even "bribes" to workers as long as these take non-pecuniary wage forms, is consistant with this. However, the standard "threat effect" approach views C_i as wages paid to nonunion workers to keep them from joining a union. It is clear, however, that even if C_i were reinterpreted as wages paid to discourage unions, for some firms, $W_U - W_N$ would overstate the effect of the union, and the average $W_U - W_N$ might also exceed $W_U - W_C$.

Let us ask a more fundamental question. If the model in this paper is a reasonable description of equilibrium in a unionized industry, what can we infer from looking at the wage differential and its relationship to other variables, especially the proportion unionized? At the risk of restating a point made by Rosen (1974) in another context, consider the following. The observed W_U, W_N pair is the outcome of solving the unions optimum problem. As one moves from left to right along a given opportunity locus, $W_U - W_N$ rises and P falls yet union "power" in the sense of opportunities stays the same.

Suppose that all unions faced the same opportunity locus and differed only on Z , the costs of running a union. Recall that as Z rises, the optimal W_U rises and P falls. So occupations for which the costs of running a union are high will have high wage differentials and few union workers. From the low P , some might infer that the union is not powerful. From the high $W_U - W_N$, others might infer that it has a great deal of power. In fact, in some sense "power" is the same across occupations because the opportunity locus is unchanged. In another sense, the high Z occupation is less powerful since its costs are higher and expected wealth is lower so that union power and proportion in the union are negatively related. Finally, a regression of $W_U - W_N$ on P will yield a negative coefficient! If the opportunity locus shifts as well across occupations, then the interpretation of the relationship between $W_U - W_N$ and the proportion in the union is even more confused. This is not to imply that studies such as the classic by Lewis (1963) or that by Rosen (1970) on wage differentials tell nothing. They describe empirical regularities which models like this one should be able to explain. But inferences drawn with respect to union power on the basis of such studies, might usefully be reexamined in light of such thinking.[11]

Quantity Restrictions and Price Discrimination:

In this seciton, we allow the union to select from a richer strategy set. Above, it was assumed that the union as monopolist could choose only the wage, W_U. However the dead weight loss which results can be eliminated and additional rent can be captured by the union if we allow price discrimination or if we allow the union to set quantity as well as price by offering all-or nothing contracts. This is illustrated in figure 4.

A monopolistic union can extract more rent than it can by simply charging the monopoly price, say $\tilde{W}_U$ and allowing the firm to hire $l' = d(\tilde{W}_U)$ workers. The union can extract up to triangle ABC in an infinite number of ways. One is to offer the firm an all-or-nothing contract to employ R/S workers (the competitive number) at wage W_U^* each (where area FCD equals area DEB). Alternatively, the union could require that the unionized firm pay a lump sum to the union equal to area ABC. The wage rate is then free to settle to the competitive level W_C, and firms voluntarily hire R/S workers. This lump sum payment is then redistributed to workers who end up earnings W_U^* as the result. The lump sum is payment to workers may take the form of fringe benefits which are paid to workers, but are invariant with respect to number of hours worked. The finding by Freeman (1978), that union workers receive a higher proportion, and therefore higher absolute amount, of compensation as fringes can be explained as part of an optimal rent extracting policy.

The advantage of the two-part wage scheme is that since the marginal wage is free to adjust, transitory changes in supply of and demand for labor are dealt with efficiently. The number of workers employed adapts automatically to the new competitive equilibrium. The all-or-nothing contract requires explicit changes in the number of workers by the union. Adjustments made by the union are more efficient if most supply or demand changes are known by the union better than the firm. If not, the two-part wage which allows the firm to select the number of workers is superior. This point, best articulated by Hall and Lilien (1979), should be coupled with another: To the extent that the lump sum transfer to the union involves the union as a intermediary, workers may prefer to avoid the possibility of skimming by the union leader through payments which come directly to workers. In a world of perfect information with respect to changes in supply and demand, this would tip the balance in favor of all-or-nothing offers. Since calling out a marginal wage necessarily implies a quantity, we conduct the following discussion in terms of the all-or-nothing offer.

It is obvious that if the firm is held captive so that it faces no alternative seller of labor, then the unions rent extracting optimum implies setting $L_U = R/S$ and wage $= W^*_U$. However, in this paper, we allow the firm to defeat the union at some cost C_i and thereby purchase labor from the nonunion firm at wage W_N. The union takes into account that higher extraction of rent implies fewer union firms from which to extract. Also since there is some probability that workers will end up in nonunion jobs, it might be preferable to select a union employment level <u>above</u> R/S and a union wage above W_C but below W_U^* so that the wage in the nonunion sector, as well as the union sector, is above W_C . (If the union firms use more than R/S workers then fewer than R/S workers are left for nonunion firms so the market clearing wage exceeds the competitive wage.)

As is proved in appendix 4, the optimum strategy for the union remains to set $L_U = R/S$. This implies that when a union organizes a firm, it will attempt to require the firm to keep the size, although not necessarily the identity, of the labor force constant at its previous level, not pushing for higher employment but not "trading off" lower employment for higher wages. It will appear, therefore, that the relevant membership is that set of workers currently employed by the newly unionized firm even though this is exactly the number of union jobs that would have been selected if all choices were made ex ante. Further, a union that can set wage and quantity brings about an efficient allocation of labor across sectors, (although not across industries or occupations).

Firms, of course, faced with paying a wage $W_U > W_C$ prefer to hire fewer than R/S workers if given the choice since the firm is off its demand curve. This has the appearance of "featherbedding," a requirement that firms hire more labor than they freely choose. Yet this featherbeding is efficient in two respects. First, it provides that a Pareto optimum is reached since the competitive number of workers are employed. Second, there is no "buy out" offer that the firm can make to the union to eliminate featherbedding which is acceptable to the union since featherbedding is an optimal rent extracting strategy.

Once we allow for quantity strategies by the union, the ambiguity in demand curve comparative statics (discussed on p. 18) disappears. Contrary to Marshall's and Hick's assumption, it turns out that elasticity of demand affects neither the probability of obtaining a union in an occupation nor the union wage if a union does exist. The intuition is this:

Given that the union fixes quantity at $L_U = R/S$, the nonunion quantity is necessarily R/S from (1) so that $W_N = W_C$. For any given wage W_U that the union calls out, the benefit from not being unionized is then $(R/S)(W_U - W_C)$. The cost of being nonunionized is C_i. Neither benefit nor cost are a funciton of the demand for labor. Therefore, the proportion of firms which resist the union is invariant with respect to the demand for labor as is the probability of observing a union in a given occupation. Similarly, since the union fixes the amount of labor hired at R/S, the probability of obtaining a union job does not vary with the demand for labor. As such, the choice of the optimum union wage does not depend upon the elasticity of demand for labor. The formal proofs are contained in appendix 5.

Whether or not the union can set quantity as well as price is open to debate. The factor that usually prevents a monopolist from price discriminating, resale of the product, does not seem to be important in this context. But both price discrimination and all-or-nothing offers give firms an incentive to change their scales of operation. For example, a union which required L_U equal to 1 1/2 times the quantity of labor that the firm would elect to purchase at price W_U might be thwarted. The firm simply would increase all other factors of production to 1 1/2 times the initial amount. In other words, lump sum requirements provide incentives for mergers. Additionally, legal restrictions, bargaining considerations, and market conditions may make it difficult for the union to set quantity as well as price. The resolution of this issue must be an empirical one.

III. An Overlapping-Generation Model

The analysis has been a static one where all workers were assumed to be identical. Yet the essence of a great deal of union activity centers on worker differences and how different preferences within the union are juggled to come out with a stable, long term, relationship. Among the most important sources of different preferences within the union are those related to the lifecycle. Old and young workers may have very different ideas about what the union should do and one wonders, for example, why such a situation does not result in one group forming its own union in competition with the other. This section proves that the multi-generational feature does not change the way in which a union behaves with respect to those issues discussed in previous sections.

What is essential about lifecycle differences within a union can be captured by a simple overlapping generation model. Assume that workers live two periods. There are Y young persons born into the occupation each period so that $R \equiv 2Y$ is the total labor supply. All workers are equally productive.

Recall that P is the proportion of workers who obtain union jobs. Before, allocation of jobs was assumed to be random. But in a two generation context the way in which jobs are assigned makes a great deal of difference. Suppose that all young workers had first claim to union jobs. If P is less than one, there will be some old workers without union jobs. They have an incentive to negotiate with the union employer, offering labor services at a wage which exceeds the nonunion wage, but is less than the union wage. The union may or may not successfully defeat such attempts at "scabbing" by older workers, but the presence of such incentives clearly raises the costs of operating the union, Z, and makes a union equilibrium less likely.

Reverse the situation. Let old workers have first claim to union jobs and the scenario is altered dramatically. Young workers, who care only about expected lifetime wealth, behave no differently. They remain loyal to the union, even as nonunion workers, because their entry into union jobs can be made contingent upon their non-disruptive behavior. Old workers, however, have no incentive to undermine the union since they are the individuals who reap the current benefits of such a scheme. As the result, no individual's ex ante lifetime wealth is reduced by a strict seniority rule for admission to the union, and the costs of operating the union are reduced, which results in a higher expected lifetime wealth for all workers. Thus, union workers will tend to be older than nonunion workers. Further, if there were retired workers present in the model, they too might desire to work occasionally and their offers of labor at less than W_U to union firms would also adversely affect expected wealth. However, if these workers can be punished for this anti-union behavior, all workers can be made better off ex ante. Pensions may play an important role here. To the extent that old workers receive large pensions which are controlled at least in part by the union, scabbing can be punished by the discontinuation of pension benefits. (A union may not be able to stop pension payments to workers, but its ability to raise the uncertainty of receiving those payments has a similar, though somewhat weaker effect.) This provides another reason why union workers should receive a greater part of their lifetime compensation in the form of pensions that nonunion workers.[12] Mincer (1981) also suggests that pensions are a way to capture rents and lessen the adverse affects of the hours reduction.

"Featherbedding" too can be rationalized along these same lines. Even featherbedding which involves inefficiency and is inferior to lump sum payoffs, has the advantage that the "bought out" individual's payment is tied to the union's continued success. A bought out worker would otherwise have an incentive to undermine the union after his payoff was received.

A word on the mechanics of this market is useful. Here, young workers play a role even though they may not be in the union. Young workers' cooperation is necessary, yet they appear as nonunion workers in this market. In some situations, workers in nonunion firms may be in the union explicitly. This is extremely common in the building trades. A worker may have his union card, but is assigned to union jobs in an order often related to seniority. If he does not obtain a union job on a given day, he may work in a nonunion job at a lower wage, but obviously cannot offer his services to the union employer. His willingness to do this rests on his knowledge that someday he will be the more senior worker and will receive W_U.

In other situations, the young worker actually is outside the union, waiting for his union card. During this period, he works nonunion jobs. Yet even this worker prefers that the occupation is unionized, because his lifetime wealth is higher as a result. Further, his cooperation is necessary in order for the union to be successful in pressing its demands.

All of this implies that the probability of being in a union job will be positively related to age in the following manner: If P_O is the probability that an old worker will be in a union and P_Y is the probability that a young worker will be in a union, then

$$P_O = \begin{cases} 1 & \text{if} \quad S[1 - G(\Pi^*(W_U,W_N))]d(W_U) > Y \\ \dfrac{S[1 - G(\Pi^*(W_U,W_N))]d(W_U)}{R - Y} & \text{if} \quad S[1-G(\Pi^*(W_U,W_N))]d(W_U) < Y \end{cases}$$

Since $Y = R/2$, this can be rewritten as

$$(19) \quad P_0 = \begin{cases} 1 & \text{if } P \geq 1/2 \\ 2P & \text{if } P < 1/2 \end{cases}$$

where $P \equiv \dfrac{S(1-G)d(W_U)}{R}$ as before .

Similarly, since young workers claim the left-over union jobs,

$$(20) \quad P_Y = \begin{cases} 2P - 1 & \text{if } P \geq 1/2 \\ 0 & \text{if } P < 1/2 \end{cases} .$$

Except when $P = 1$, $P_0 > P_Y$ so that average age of union workers will be higher than that for nonunion workers. In this simple framework, if $\bar{A}_0$ is the average age of all old workers and $\bar{A}_Y$ is the average age of young workers (where old and young refer to their priority levels for union jobs, all being the same within the class), then the average age of union workers, $\bar{A}_U$, can be obtained. Since there are $(P)(R)$ union workers, $(P_0)(Y)$ old workers and $(P_Y)(Y)$ young workers.

$$\bar{A}_U = \frac{P_0 Y}{PR} \bar{A}_0 + \frac{P_Y Y}{PR} \bar{A}_Y$$

or

$$(21) \quad \bar{A}_U = \frac{P_0}{2P} \bar{A}_0 + \frac{P_Y}{2P} \bar{A}_Y .$$

Analogously, for nonunion workers

$$\bar{A}_N = \frac{Y(1-P_0)}{(1-P)R}\bar{A}_0 + \frac{Y(1-P_Y)}{(1-P)R}\bar{A}_Y$$

or

$$(22) \qquad \bar{A}_N = \frac{(1-P_0)}{2(1-P)}\bar{A}_0 + \frac{(1-P_Y)}{2(1-P)}\bar{A}_Y .$$

These simple formulas yield a number of testable empirical implications. First substituting (19) and (20) into (21), (21) can be written as

$$(23) \quad \begin{aligned} &\text{a.} \quad \bar{A}_U = \frac{P_0}{2P} A_0 = \bar{A}_0 \quad \text{if} \quad P < 1/2 . \\ &\text{b.} \quad \bar{A}_U = \frac{\bar{A}_0}{2P} + \frac{(2P-1)}{2P}\bar{A}_Y \quad \text{if} \quad P \geqslant 1/2 . \end{aligned}$$

Similarly, (22) can be rewritten as

$$(24) \quad \begin{aligned} &\text{a.} \quad \bar{A}_N = \frac{(1-2P)\bar{A}_0}{2(1-P)} + \frac{\bar{A}_Y}{2(1-P)} \quad \text{if} \quad P < 1/2 \\ &\text{b.} \quad \bar{A}_N = \frac{2-2P}{2(1-P)}\bar{A}_Y = \bar{A}_Y \quad \text{if} \quad P \geqslant 1/2 . \end{aligned}$$

Differentiating (23) with respect to P yields:

$$(25) \quad \begin{aligned} &\text{a.} \quad \frac{\partial \bar{A}_U}{\partial P} = 0 \quad \text{if} \quad P < 1/2 \\ &\text{b.} \quad \frac{\partial \bar{A}_U}{\partial P} = \frac{\bar{A}_Y - \bar{A}_0}{2P^2} < 0 \quad \text{if} \quad P \geqslant 1/2 . \end{aligned}$$

Similarly, differentiating (24) yields

$$(26)\quad \begin{aligned} &\text{a.} \quad \frac{\partial \bar{A}_N}{\partial P} = \frac{\bar{A}_Y - \bar{A}_0}{2(1-P)^2} < 0 \quad \text{if} \quad P < 1/2 \\ &\text{b.} \quad \frac{\partial \bar{A}_N}{\partial P} = 0 \quad \text{if} \quad P \geq 1/2 \ . \end{aligned}$$

The sense of (25) and (26) is this: If $P < 1/2$, then no young workers are in the union and some old workers are forced to work in the nonunion sector. Raising P simply brings more old workers into the union, but does not change the average age there since all union workers were and remain old. Therefore $\partial \bar{A}_U / \partial P = 0$ when $P < 1/2$. The nonunion pool has all of the young workers plus some of the old. As P rises, old workers are drawn from the nonunion sector into the union sector, leaving a larger proportion of young workers left in the nonunion jobs and thereby lowering $\bar{A}_N$. Therefore $\partial \bar{A}_N / \partial P < 0$.

Alternatively, if $P \geq 1/2$, all old workers, some young ones as well, are in the union. An increase in P brings more young workers into the union thereby lowering $\bar{A}_U$. At the same time fewer workers remain in the nonunion sector, but they remain as before, only young workers. Therefore $\bar{A}_N$ does not change.

This yields an empirically testable implication. A regression of $\bar{A}_U$ on P should yield a negative coefficient as should a regression of $\bar{A}_N$ on P .[13] Further, by using (23b) and (24a) a regression which defines as observations occupations for which $P \geq 1/2$ when the relationship is (23b) and $P < 1/2$ when the relationship is (24a), the coefficients of the pooled regression of $\bar{A}_U$ on $1/2P$ and $\frac{2P-1}{2P}$ for (23b) and of $\bar{A}_N$ on $\frac{1-2P}{2(1-P)}$ and $\frac{1}{2(1-P)}$ for (24a) yields the estimates of $\bar{A}_0$ and $\bar{A}_Y$. Since "old" workers is defined in terms of ages over which priority on entrance to the union is the same, and similarly for "young," this is a summary statistic on the age stratification of unions. For example $\hat{\bar{A}}_0 = \hat{\bar{A}}_Y$ implies that age is not a criterion for union membership.

It is interesting to ask how the difference between $\bar{A}_U$ and $\bar{A}_N$ varies with P. Note that from (23) and (24), one can write

$$(27)\quad \bar{A}_U - \bar{A}_N = \begin{cases} (\bar{A}_0 - \bar{A}_Y)/2(1-P) & \text{for} \quad P < 1/2 \\ (\bar{A}_0 - \bar{A}_Y)/2P & \text{for} \quad P \geq 1/2 \end{cases}$$

so that

$$(28)\quad \frac{\partial(\bar{A}_U - \bar{A}_N)}{\partial P} = \begin{cases} (\bar{A}_0 - \bar{A}_Y)/2(1-P)^2 > 0 & \text{for} \quad P < 1/2 \\ (\bar{A}_Y - \bar{A}_0)/2P^2 < 0 & \text{for} \quad P \geq 1/2 . \end{cases}$$

Since $\bar{A}_Y < \bar{A}_0$, the difference between the average age of union workers and that of nonunion workers within an occupation will first rise, then fall with P. This, too, is easily verified empirically. Also note that

$$\lim_{P\to 0} \bar{A}_U - \bar{A}_N = \lim_{P\to 0} \bar{A}_U - \bar{A}_N = \frac{\bar{A}_0 - \bar{A}_Y}{2} .$$

It is obvious from (23) and (24) that

$$\frac{\partial \bar{A}_U}{\partial \bar{A}_0} , \frac{\partial \bar{A}_U}{\partial \bar{A}_Y} , \frac{\partial \bar{A}_N}{\partial \bar{A}_0} , \frac{\partial \bar{A}_N}{\partial \bar{A}_Y} > 0$$

so that an increase in the average age of the relevant labor pool will increase the average age of both union and nonunion workers. However, for a given P , an increase in $\bar{A}_0$ and $\bar{A}_Y$ which is neutral in the sense that it leaves $\bar{A}_0-\bar{A}_Y$ unchange will also leave $\bar{A}_U-\bar{A}_N$ unchanged. This follows directly from (27). These relationships are also empirically verifiable in the sense that they give definite predictions on the relationship between the age structure of an occupation and the age structure between union and nonunion workers within that occupation.[14]

Attrition:

Some occupations are characterized by a more fickle labor force than others. Occupations with high turnover or attrition rates present additional difficulties for the prospective union. First, such movement makes it more difficult to keep tabs on which workers "paid their dues" when young by accepting a nonunion job without attempting to bargain away a union job from a more senior worker. This factor by itself raises Z , the costs of operating a union. As such, it will reduce the likelihood of a union equilibrium, but will raise the wage differential and lower the proportion of union workers in those occupations where a union equilibrium is obtained. This suggest that occupations such as secretaries and farm workers where mobility into and out of the labor force or between geographical regions is high are less likely to be unionized. It also suggests that in high attrition occupations where unions are formed, e.g., the California farm workers, only a small proportion of workers will be employed by unionized firms and the union-nonunion wage differential will be large.

Additionally, increased attrition, to the extent that it reflects ex ante known individual differences, makes a union equilibrium less likely even if it does not affect Z . The reason is that an individual who plans to leave the occupation between period one and two receives an expected wage rate in period 1 of $(W_U - Z)(P_Y) + W_N(1 - P_Y)$. He favors the union if

$$(29) \quad W_C < (W_U - Z)P_Y + W_N(1 - P_Y) \; .$$

This is a stronger condition that the one relevant for those who plan to remain in the occupation for their entire lifetimes. That condition is

$$(30) \quad 2W_C < (W_U - Z)(P_Y + P_0) + W_N[(1 - P_Y) + (1 - P_0)] \; .$$

It is easy to see that (29) is sufficient for (30), but that (30) can hold when (29) does not, so that young workers who plan to leave the occupation are more likely to oppose the certification of a union.[15] Further, attrition of workers between period one and two increases P_0 for a given labor force, R . This makes the stayers even more anxious for a union equilibrium. Since NLRB rules give each worker one and only one vote, strong differences in lifetime plans across workers reduces the probability of acquiring a union equilibrium.

A corollary of the previous discussion is that old workers will prefer unions before young workers do as long as their seniority is granfathered. For old workers to prefer the union to a competitive equilibrium, we require

It is easy to show that (30), the condition for the young to favor a union, is a sufficient condition for (31).[16]

The Relationship Between the Multi-Period Model and the One-Period Model Results:

The link between the overlapping generation model and the simple period model is this: Young workers in the multiperiod world will behave exactly like single period wealth maximizers. Their "critical indifference curves" and their view of the opportunity locus is identical to that of individuals who live only a single period.[17] However, old workers may prefer a union equilibrium even when single period workers prefer the competitive equilibrium because $P_0 \geq P$. This also implies a preference by old workers for a union wage which deviates from the solutions obtained in the one-period model.

This poses no problem under a number of circumstances. The most straightforward resolution relies on the fact that the median voter is a young worker.[18] Although all young workers are not formal members of the union, their explicit or implicit acquiescence is necessary for the maintenance of a union equilibrium. Old workers would like to ignore the nonunion young workers in selecting W_U, but their persistence in following that strategy is not viable because young workers have an incentive to undermine such a union and replace it with a lifetime wealth maximizing one. In fact, the young could pay old workers enough to induce the old workers to follow the broader lifetime wealth maximizing policy. Entry fees and contributions by the current union workers to the already retired workers' pension fund might be reinterpreted as a transfer of this sort.

Since young workers behave the same as workers who live for only a single period, and since a young worker is the marginal worker, the solution, which reflects the preference of the young worker, is identical to that in the single period world. Therefore, all results of that section hold.

Age Earnings Profiles:

Given the way that the union is structured so that young workers are less likely to be unionized than are old workers, there is a natural tendency for wages to grow over the lifecycle as workers move from nonunion to union jobs. This is true even if productivity does not grow over the lifecycle.

More important, is that within a union firm, age-earnings profiles will be flatter than they are in nonunion firms. The reason is that unions, in order to maintain stability, want the old rather than young workers to receive the benefits of the union. But in most industries and occupations, the firm, rather than the union, hires workers. A way by which the union can discourage the firm from hiring young workers is to overprice them relative to more experienced workers. This will provide the firm with incentives to hire the workers into the union firm which the union would have selected. Thus, the union can implicitly control hiring and the compensation of the firms' work force simply by calling out the appropriate wages.

A corollary is that if age and skill are positively correlated, higher quality labor will be found in union firms. Union firms, faced with flatter age-earnings voluntarily select the older highly skilled workers, because they are underpriced relative to the younger, less skilled workers. An alternative is provided by Mincer (1981) who cites evidence that less on-the-job training occurs in union firms. There is no obvious theoretical reason why this should be the case.

Incidentally, if the bimodal age distribution is replaced by a continuous one, nothing is altered fundamentally but P is reinterpreted as the proportion of one's life spent in the union. Incentives remain intact and all conclusions still follow. (Of course, the formulas for the average ages of workers in union and nonunion jobs would require alteration.)

Countercyclical Variations in Wage Differentials:

Lewis reports that union workers do relatively better during cyclical downturns. The age-based incentive mechanism provides an explanation. Since incentive compatibility requires that old workers in unions receive benefits relative to young, we expect that seniority will be more important in determining layoff priority in union firms. As such, the average age of union workers relative to nonunion workers should rise during cyclical downturns and since age-earnings profiles are positively sloped, wage differentials will increase. This is easily tested by examining the way in which $\bar{A}_U - \bar{A}_N$ moves over the business cycle. Also, controlling for age should eliminate most of the countercyclical wage differential movement.

Definition of Variables

S	Number of Firms
W	Wage rate
L	Number of workers in the firm
$R(W)$, R	Supply of labor
$d(W)$	Demand for labor per firm
W_U	Union wage
W_N	Nonunion wage
C_i	Cost of fighting union for firm i
$g(C)$	Density function
$G(C)$	Distribution function
$\Pi(W)$	Profit function
$\Pi^*(W_U, W_N)$	$\Pi(W_N) - \Pi(W_U)$
W_C	Competitive wage in the absence of unions
$dW_N/dW_U \mid_{OL}$	Slope of opportunity locus
P	Probability of obtaining a union job
$dW_N/dW_U \mid_{IC}$	Slope of indifference curve
Z	Per member cost of operating the union
Critical indifference curve	Indifference curve that yields same level of utility as available if the worker receives competitive wage, W_C.
$f(1)$	Production function of firm
$D(p_Q)$	Demand for product as function of price P_Q
Y	Number of young entrants to an occupation
P_0	Probability that an old worker is in a union

P_Y	Probability that a young worker is in a union
$\bar{A}$	Average age of all workers
$\bar{A}_O$	Average age of old workers
$\bar{A}_Y$	Average age of young workers
$\bar{A}_U$	Average age of union workers
$\bar{A}_N$	Average age of nonunion workers

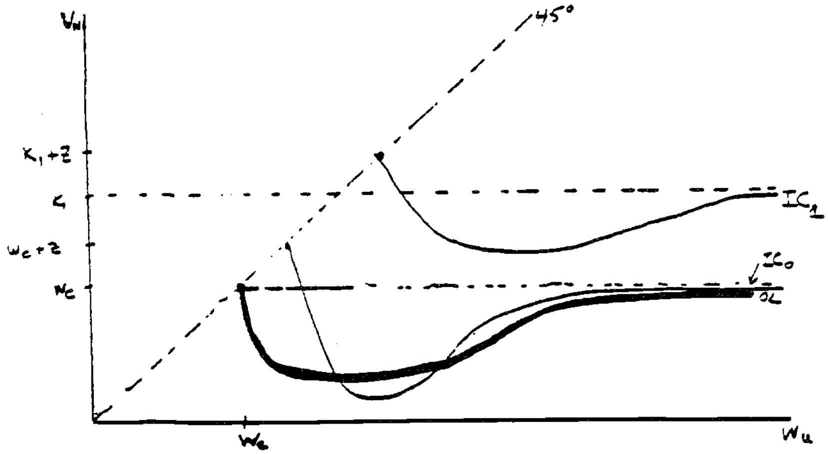

Figure 1

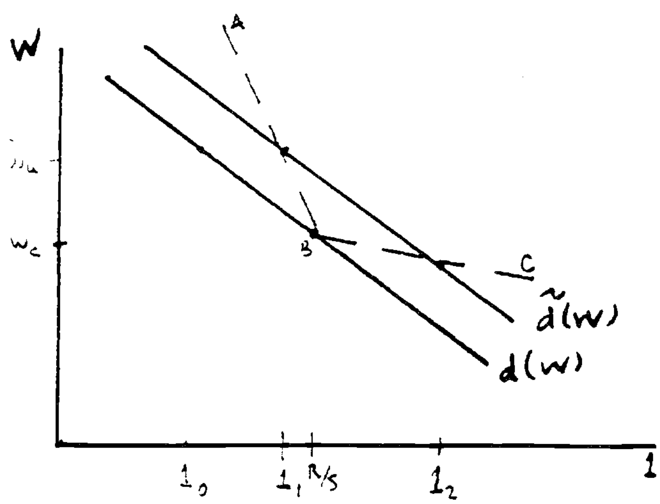

Figure 2

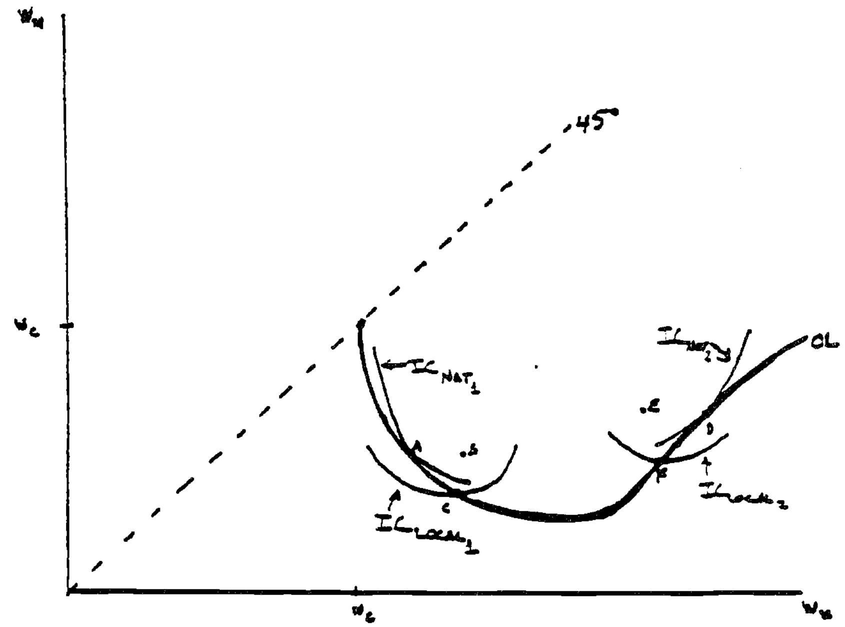

Figure 3

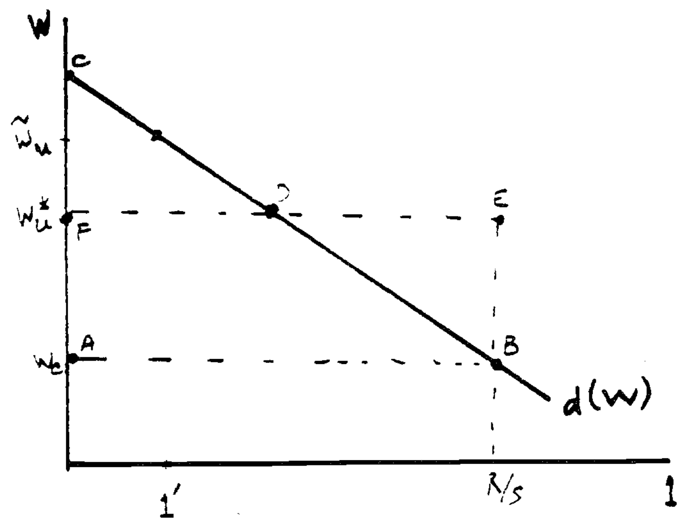

Figure 4

Appendices

1 : Proof that elastic labor supply shifts the opportunity locus downward.

By contradiction: Define $\tilde{W}_N$ as W_N corresponding to $R(K)$ and W_N corresponding to $R(W_C) \equiv R$. Assume that $\tilde{W}_N > W_N$. Given that $R(K) > R$, we know that

$$(1-G(\Pi^*[W_U,\tilde{W}_N]))d(W_U)+G(\Pi^*[W_U,\tilde{W}_N])d(\tilde{W}_N) > [1-G(\Pi^*[W_U,W_N])]d(W_U) + G(\Pi^*[W_U,W_N])d(W_N)$$

But if $\tilde{W}_N > W_N$ then $d(W_N) > d(\tilde{W}_N)$ so

$$(1-\tilde{G})d(W_U)+\tilde{G}d(W_N) > (1-G)d(W_U)+Gd(W_N)$$

or $d(W_U)(G-\tilde{G}) > d(W_N)(G-\tilde{G})$.

Since $W_U > W_N$, this implies $\tilde{G} > G$. But if $\tilde{W}_N > W_N$ $\tilde{G} < G$ since $\Pi_2^* < 0$ and $G' = g > 0$. This results in a contradiction.

Proof that critical indifference curve shifts upward:

Define $\tilde{P}$ as the probability of obtaining a union job if $R = R(K)$. Differentiating (7) with respect to P yields $\left.\frac{dW_N}{dP}\right|_{IC} = \frac{K+Z-W_U}{(1-P)^2} < 0$ since $W_U > K+Z$ for a union equilibrium to exist. Therefore, if $\tilde{P} < P$ then $\tilde{W}_N > W_N$ and the indifference curve shifts up.

Assume $\tilde{P} > P$. Then $\frac{S}{R(K)}(1-\tilde{G})d(W_U) > \frac{S}{R}(1-G)d(W_U)$. Since $R(K) > R$, this implies that $1-\tilde{G} > 1-G$ or that $\tilde{G} < G$. But if $\tilde{P} > P$ then $\tilde{W}_N < W_N$ since $dW_N/dP < 0$. Since $\Pi_2^* < 0$, this implies $\tilde{G} > G$ which is a contradiction. Therefore the critical indifference curve shifts upward.

2: Proof that a parallel shift in demand affects neither the probability of a union equilibrium nor the wage differential selected.

It is sufficient to show that the opportunity locus and relevant indifference curves are displaced along a 45° ray. If so, all solutions will move in proportion to the new competitive wage.

Consider $\tilde{d}(W) > d(W)$ such that $\tilde{d}'(W)=d'(W)=d' \;\forall\; W$. Define $\tilde{W}_C$ such that $\tilde{d}(\tilde{W}_C)=R/S$ and $\Delta \equiv \tilde{W}_C - W_C$. Then, if $\tilde{W}_U, \tilde{W}_N$ are pairs on the new opportunity locus, a parallel shift of that locus requires that $\tilde{W}_N=W_N+\Delta$ if $\tilde{W}_U=W_U+\Delta$.

The equation for the opportunity locus (eq. (1)) implies that at $\tilde{W}_U=W_U+\Delta$,

$$\text{(A2.1)} \qquad [1-G(\Pi^*(W_U,W_N))]d(W_U)+G(\Pi^*(W_U,W_N))d(W_N)=R/S$$
$$= [1-G(\tilde{\Pi}^*(W_N+\Delta,\tilde{W}_U)]\tilde{d}(W_U+\Delta)+G(\tilde{\Pi}^*(W_U+\Delta,\tilde{W}_N))\tilde{d}(\tilde{W}_N)$$

Conjecture that

$$\Pi^*(W_U,W_N) = \tilde{\Pi}^*(W_U+\Delta,\tilde{W}_N) \ .$$

Then (A2.1) implies that

$$d(W_N) = \tilde{d}(\tilde{W}_N) \ .$$

or

$$d(W_C) + (d')(W_N-W_C) = \tilde{d}(W_C) + (\tilde{d}')(\tilde{W}_N - \tilde{W}_C) \ .$$

Since $d(W_C) = \tilde{d}(\tilde{W}_C)$, $d' = \tilde{d}'$, and $\tilde{W}_C - W_C = \Delta$, this implies that $\tilde{W}_N = W_N + \Delta$. But if $\tilde{W}_N = W_N + \Delta$ when $W_U = W_U + \Delta$, then since $\tilde{d}(W_U + \Delta) = d(W_U)$ and $\tilde{d}(W_N + \Delta) = d(W_N)$, $\Pi^*(W_U,W_N) = \tilde{\Pi}^*(W_U+\Delta,W_N+\Delta)$ so $G=\tilde{G}$ and the sufficient condition is verified. Therefore a parallel shift in the opportunity locus occurs.

Similarly, the critical indifference curve is displaced up the 45° line. Using (12),

$$\text{(A2.2)} \qquad \tilde{W}_N = \frac{\tilde{W}_C - \tilde{P}(\tilde{W}_U - Z)}{1 - \tilde{P}} \quad .$$

Evaluating this at $\tilde{W}_U = W_U + \Delta$ and recalling that, for all points on the opportunity locus, $G = \tilde{G}$, (A2.2) can be rewritten as

$$\tilde{W}_N = \frac{W_C + \Delta - P(W_U + \Delta - Z)}{1 - P}$$

$$= \frac{W_C - P(W_U - Z)}{1 - P} + \Delta$$

$$= W_N + \Delta \qquad \text{(from 12)}.$$

So the critical indifference curve is displaced along the 45° line. This implies that the probability of obtaining a union equilibrium does not vary when linear demand for labor shifts out parallel.

Finally, the equilibrium wage differential does not change because the slope of the new indifference curves at the new opportunity locus exactly equals the slope of the old indifference curves at the old opportunity locus along the 45° ray. Using (8), (9), and (10),

$$\text{(A2.3)} \qquad \left.\frac{d(W_N+\Delta)}{d(W_U+\Delta)}\right|_{\tilde{IC}} = - \frac{1-G(\tilde{\Pi}^*(W_U+\Delta,W_N+\Delta))+(W_U+\Delta-Z-W_N-\Delta)[(1-G(\tilde{\Pi}^*(W_U+\Delta,W_N+\Delta)))\frac{\tilde{d}'(W_U+\Delta)}{\tilde{d}(W_U+\Delta)} - g(\tilde{\Pi}^*(W_U+\Delta,W_N+\Delta))\tilde{d}(W_U+\Delta)]}{\frac{R}{S} - (1-G(\tilde{\Pi}^*(W_U+\Delta,W_N+\Delta)))+(W_U+\Delta-Z-W_N-\Delta)g(\tilde{\Pi}^*(W_U+\Delta,W_N+\Delta))\tilde{d}(W_N+\Delta)}$$

But since $\tilde{d}' = d'$, $\tilde{\Pi}^*(W_U+\Delta,W_N+\Delta) = \Pi^*(W_U,W_N)$, $\tilde{d}(W_U+\Delta) = d(W_U)$ AND $\tilde{d}(W_N+\Delta) = d(W_N)$, (A2.3) can be rewritten as

$$\left.\frac{d(W_N+\Delta)}{d(W_U+\Delta)}\right|_{\tilde{IC}} = - \frac{1-G(\Pi^*(W_U,W_N))+(W_U-Z-W_N)[(1-G(\Pi^*(W_U,W_N)))\frac{d'(W_U)}{d(W_U)} - g(\Pi^*(W_U,W_N))d(W_U)]}{\frac{R}{S} - (1-G(\Pi^*(W_U,W_N)))+(W_U-Z-W_N)g(\Pi^*(W_U,W_N))d(W_N)}$$

$$= \left.\frac{dW_N}{dW_U}\right|_{IC}$$

Proof that the more convex the demand curve, the more likely is a union equilibrium to exist:

The formal proposition: Suppose a convex demand curve, $\tilde{d}(W)$, is tangent to a linear demand curve $d(W)$ at the competitive equilibrium $[R/S, W_C]$. If a union equilibrium exists for $\tilde{d}(W)$, then it exists for $d(W)$ although the converse is not true.

First, the opportunity locus for $\tilde{d}(W)$ lies above that for $d(W)$ (except at $W_U = W_C$). To see this, assume the opposite. Then for every given W_U , $W_N > \tilde{W}_N$. From (1),

$$(A2.4) \qquad [1-G(\tilde{\Pi}^* W_U,\tilde{W}_N))]\tilde{d}(W_U)+G(\tilde{\Pi}^*(W_U,\tilde{W}_N))\tilde{d}(\tilde{W}_N)$$

$$= R/S = [1-G(\Pi^*(W_U,W_N))]d(W_U)+G(\Pi^*(W_U,W_N))d(W_N)$$

But if $W_N > \tilde{W}_N$, then $d(W_N) < d(\tilde{W}_N) < \tilde{d}(\tilde{W}_N)$ and then $G(\tilde{\Pi}^*(W_U,\tilde{W}_N)) > G(\Pi^*(W_U,W_N))$ therefore

$$G(\tilde{\Pi}^*(W_U,\tilde{W}_N))\tilde{d}(\tilde{W}_N) > G(\Pi^*(W_U,W_N))d(W_N) .$$

This, along with (A2.4), implies that

$$[1-G(\tilde{\Pi}^*(W_U,\tilde{W}_N))]\tilde{d}(W_U) < [1-G(\Pi^*(W_U,W_N))]d(W_U)$$

or that $1 < \tilde{G} - G$ which is a contradiction since $0 \leq G,\tilde{G} \leq 1$.

Also, the critical indifference curve for $\tilde{d}(W)$ lies below that for $d(W)$: Assume the opposite, $\tilde{W}_N > W_N$. Then from A1, $\tilde{P} < P$. But $\tilde{d}(W_U)>d(W_U)$ and $\tilde{W}_N > W_N$ implies that $(1-\tilde{G})<(1-G)$ or $\tilde{G} > G$ which implies $\tilde{W}_N < W_N$ which is a contradiction.

3: As the cost of running the union rises, the probability of obtaining a union job tends to fall: From (4),

$$dP = \frac{-S}{R}g(\ )\ [\Pi_1^* dW_U + \Pi_2^* dW_U] + \frac{S}{R}[1-G(\ )]d'(W_U)dW_U$$

$$= \frac{-S}{R}g(\ )\ [d(W_U)dW_U - d(W_N)dW_N] + \frac{S}{R}[1-G(\ )]d'(W_U)dW_U\ .$$

where $dW_N = \left.\frac{dW_N}{dW_U}\right|_{OL} dW_U$. The second term is always negative. For much of the opportunity locus $dW_U > 0$ implies $dW_N < 0$. For the part where $dW_N > 0$, it is smaller than dW_U so dP tends to be negative. Thus, an increase in operating costs lowers the likelihood of a union equilibrium, but if one does exist, raises the optimal union wage, wage differential, and tends to lower the proportion in the union.

4: Proof that a union which can choose price and quantity sets quantity equal to R/S (even though this affects the wage in the nonunion sector and the probability of being beaten by the firm) and this results in $W_N = W_C \; \forall \; W_U$;

Define l_U and l_N as the labor per firm in the union and nonunion sector. Then the firm, presented with wage-quantity demand (W_U , l_U) compares the profits associated with it against that of (W_N, l_N). If this difference is smaller than C_i the firm fights (and defeats) the union. The profit function now depends upon quantity as well as price and we define

$$\Delta\Pi(W_U, W_N; l_U, l_N) \equiv \Pi(W_N; l_N) - \Pi(W_U; l_U)$$

(A4.1)
$$= \left\{ \int_0^{l_N} [d^{-1}(l)]dl - l_N W_N - \text{fixed cost} \right\} - \left\{ \int_0^{l_U} [d^{-1}(l)]dl - l_U W_U - \text{fixed cost} \right\}$$

$$= \int_{l_U}^{l_N} [d^{-1}(l)]dl + l_U W_U - l_N W_N .$$

For any $[W_U, W_N; l_U, l_N]$ there will be $1-G[\Delta\Pi(W_U, W_N; l_U, l_N)]$ union firms. The unions problem, then, is to select W_U, l_U and implicitly W_N, l_N so as to maximize

(A4.2) $\quad P(W_U - Z) + (1-P)W_N$

subject to the constraint that

(A4.3) $$[1-G[\Delta\Pi(W_U,W_N;l_U,l_N)]l_U+G(\quad)l_N = R/S$$

where

(A4.4) $$P = \frac{S}{R}(1-G)l_U \ .$$

Forming the Lagrangean:

(A4.5) $$L = P(W_U-Z)+(1-P)W_N+\lambda\{[1-G(\quad)]l_U+G(\quad)l_N-R/S\} \ .$$

Since

(A4.6) $$\frac{\partial\Delta\Pi}{\partial W_U} = l_U; \ \frac{\partial\Delta\Pi}{\partial W_U} = -l_N \ ;$$
$$\frac{\partial\Delta\Pi}{\partial l_U} = W_U-d^{-1}(l_U) \ ;$$
$$\frac{\partial\Delta\Pi}{\partial l_N} = d^{-1}(l_N)-W_N \ ,$$

The first order conditions for (A4.5) are

(A4.7)

$$\text{a.}\quad \frac{\partial L}{\partial W_U} = (W_U - Z - W_N)\frac{\partial P}{\partial W_U} + P + \lambda[(l_N - l_U)(G')l_U] = 0$$

$$\text{b.}\quad \frac{\partial L}{\partial W_N} = (W_U - Z - W_N)\frac{\partial P}{\partial W_N} - P - \lambda[(l_N - l_U)(G')l_N] = 0$$

$$\text{c.}\quad \frac{\partial L}{\partial l_U} = (W_U - Z - W_N)\frac{\partial P}{\partial l_U} + \lambda[1 - G + (l_N - l_U)(G')(W_U - d^{-1}(l_U))] = 0$$

$$\text{d.}\quad \frac{\partial L}{\partial l_N} = (W_U - Z - W_N)\frac{\partial P}{\partial l_N} + \lambda[G + (l_N - l_U)(G')(d^{-1}(l_N) - W_N)] = 0$$

$$\text{e.}\quad (1 - G(\quad))l_U + G(\quad)l_N - R/S = 0\ .$$

Note from (A4.4) that

(A4.8)

$$\frac{\partial P}{\partial W_U} = \frac{S}{R}\, l_U(G')l_U$$

and that

$$\frac{\partial P}{\partial W_N} = \frac{S}{R}\, l_U(G')l_N\ .$$

Rearranging (A4.7 a,b) and dividing (A4.7a) by (A4.7b) gives

(A4.9) $\quad l_U = l_N\ .$

Using (A4.7e) implies that $l_U = l_N = R/S$.

Also since $d(W_N) = R/S$, $W_N = W_C$.

5: Proof that more inelastic demand changes neither the probability of a union or the nonunion wage where unions cna select price and quantity.

Consider two labor demand curves, $d(W)$ and $\tilde{d}(W)$ such that $d(W_C) = \tilde{d}(W_C) = R/S$ but $|d'| < |\tilde{d}'|$ as shown. Given that $l_N = l_U = R/S$, $W_N = W_C$ (see appendix 4), the opportunity locus is a horizontal line at $W_N = W_C$. Therefore, to show that the probability of a union does not change when going from $d(W)$ to $\tilde{d}(W)$ it is sufficient to show that the indifference curves do not shift:

The equations of a indifference curve producing utility level K given demand for labor $d(W)$ and $\tilde{d}(W)$, respectively are

$$\text{(A5.1)} \quad \text{a.} \quad W_N = \frac{K - P(W_U - Z)}{(1-P)}$$

$$\text{b.} \quad \tilde{W}_N = \frac{K - \tilde{P}(W_U - Z)}{(1-\tilde{P})}$$

where

$$P = \frac{S(1-G(\Delta\Pi))R/S}{R} = 1-G(\Delta\Pi)$$

and

$$\tilde{P} = 1-G(\tilde{\Delta\Pi}) .$$

Since $l_U = l_N = R/S$, $\Delta\Pi$ (defined in appendix 4) is $(R/S)(W_U - W_N)$ so that (A5.1a,b) becomes

(A5.2)

$$\text{a.} \quad W_N = \frac{K-G[(R/S)(W_U-W_N)](W_U-Z)}{G[(R/S)(W_U-W_N)]}$$

$$\text{b.} \quad \tilde{W}_N = \frac{K-G[(R/S)(\tilde{W}_U-W_N)](W_U-Z)}{G[(R/S)(W_U-W_N)]}$$

Substitution of (A5.2a) into (A5.2b) yields $W_N = \tilde{W}_N$ so indifference curves are identical.

Corollary: Since indifference curves and opportunity locus are invariant with respect to $d(W)$, $\tilde{d}(W)$, it follows that optimal W_U is the same in both cases.

FOOTNOTES

The author gratefully acknowledges the helpful comments of K. Abraham, G. Becker, R. Hall, V. Lazear, J. Mincer, G. Neumann, P. Pashigian, M. Reder, S. Rosen and R. Topel.

[1]The relevant supply of labor is that pool over which employers view workers as perfect substitutes in production. This is narrower than Ross's (1948) "orbit of coercive comparison," or Dunlop's (1944) "wage contours."

[2]It is simplifying, but inessential, to assume that $g(C_i)$ does not depend upon $W_U - W_N$ directly. Workers might strive harder for the union if $W_U - W_N$, or more exactly the expected utility gain, is larger. It is assumed that $G(C_i)$ is known, by all, but the particular C_i is known only to the firm.

[3]This subsumes all bargaining problems. Thus, Reder's (1952) notions of fairness and Steven's (1958) early description of bargaining, as well as more modern bargaining models (Farber (1978), Crawford (1979), Atherton (1973)) are implied.

[4]Johnson and Miezkowski (1970) and Diewert (1974) offer alterntive two-sector models of unionism. There, the union wage is exogenous and the purpose is to trace various wage choices through the rest of the economy in a general equilibrium framework. This model is interested in the way in which spillovers affect the choice of a union wage.

[5]This is a major break with the literature in that most of what has gone before, attempts to limit the relevant population to some subset of R. For example, Reder (1959) focuses on "present members." But present at what point in time? How does the union ever get started under these assumptions? Also, Dertouzos and Pencavel estimate a wage and employment relationship using data

from the International Typographical Union. This is not equivalent to my indifference curve, however, because it abstracts from spillover effects of W_U on W_N and changes in the opportunity locus over time.

[6]Even in the pathological case where $g(0)$, Z, and $d(W_U)$ are sufficiently large so that the indifference curve is initially positive, it will rapidly become negative because when $W_U = W_N + Z$,

$$\left.\frac{dW_N}{dW_U}\right|_{IC} = \frac{-P}{1-P} < 0 .$$

[7]Stiglitz (1980) obtains a similar result in a different context. He shows that convexity affects whether a random taxation scheme dominates a nonrandom one.

[8]In this context, Epple, Hotz and Zelenitz (1980) formalize the argument that in high variance demand industries, the union performs a risk pooling function which reduces the necessity of formal layoffs. The union acts as the hiring hall and assigns workers accordingly. Fluctuation in an individual firm's demand does not result in a "layoff," as the result.

[9]There are additional second-order effects. There is no guarantee that the reduction in output by union firms plus the increase in output by nonunion firms will yield the net reduction in total output required as we move up the product demand curve. This will require a change in S, the number of firms in the industry.

[10]Rosen (1969) treats this issues.

[11]Recent work by Mincer (1981) makes a similar point, but exploits a different mechanism. Mincer points out that altering the probability of obtaining a job in the union sector affects the expected return to queuing for jobs in that sector and affects spillovers to the nonunion sector. Depending

upon the nature of these spillovers, the observed wage differential may overstate the true effect of unions. He, too, concludes that there is no straightforward relationship between wage differentials and power. The main difference between his approach and this one is that in this model queuing for union jobs is done while holding a nonunion job so there is never any unemployment. Mincer's queuing takes place while the worker is unemployed.

[12]See Freemen (1978) for evidence in support of this prediction.

[13]This assumes that $\bar{A}_0 - \bar{A}_Y$ does not vary across occupations.

[14]In principle, there could be a larger number of age groups than two, and one extreme specification would allow a group for each age level, say measured in years, so that $\bar{A}_0$ and $\bar{A}_Y$ would be replaced by $A_{18}, A_{19} \ldots A_{65}$.

[15]Proof:

$$W_C < (W_U - Z)P_Y + W_N(1 - P_Y)$$

implies that

$$2W_C < (W_U - Z)2P_Y + W_N(2)(1 - P_Y)$$

Since $(W_U - Z) > W_N$ the convex combination $(W_U - Z)\lambda + W_N(1 - \lambda)$ increases in λ. Since $P_Y + P_0 > 2P_Y$ it follows that $2W_C < (W_U - Z)2P_Y + W_N(2)(1 - P_Y) < (W_U - Z)(P_Y + P_0) + W_N[(1 - P_Y) + (1 - P_0)]$, so (29) is sufficient for (30).

[16]Proof:

Equation (30) is

$$2W_C < (W_U - Z)(P_Y + P_0) + W_N[(1 - P_Y) + (1 - P_0)] .$$

This implies

$$W_C < (W_U - Z)(P_Y + P_0)/2 + W_N[(1 - P_Y) + (1 - P_0)]/2$$

The r.h.s. is a convex combination of $(W_U - Z)$, W_N with $(W_U - Z) > W_N$. Now since

$$P_0 \geq P_Y , \; P_0 \geq (P_0 + P_Y)/2$$

therefore

$$(W_U-Z)P_0+W_N(1-P_0)\geq(W_U-Z)[(P_0+P_Y)/2]+W_N[(1-P_0)+(1-P_Y)]/2 > W_C$$

so (30) is sufficient for (31).

[17]Proof:

Young vote for the union if (30) holds, i.e., if

$$2W_C < (W_U-Z)(P_Y + P_0)+W_N[(1-P_Y)+(1-P_0)]$$

substituting in (19) and (20) this can be rewritten as

$$2W_C < (W_U-Z)(2P)+W_N[2(1-P)]$$

or

$$W_C < (W_U-Z)P+W_N(1-P)$$

which is the condition for a single-period lived worker to prefer the union equilibrium (derivable from equation (12)).

[18]See Farber (1978) for discussion of some basic aspects of union equilibrium in a median voter world.

REFERENCES

Atherton, W. Theory of Union Bargaining Goals, Princeton: Princeton University Press 1973.

Blau, Francine and Lawrence Kahn. "The Exit-Voice Model of Unionism: Some Further Evidence on Layoffs," University of Illinois, March 1980.

Brown, Charles and James Medoff. "Trade Unions in the Production Process." Journal of Political Economy, 86,3, pp. 355-78.

Carlton, Dennis. "Planning and Market Structure," in The Economics of Uncertainty, ed. by J.J. McCall (1979).

Cohen, K.J. and R.M. Cyert. Theory of the Firm: Resource Allocation in a Market Economy, 2nd edition, Prentice Hall: Englewood Cliffs, NJ (1975).

Crawford, Vincent P. "On Compulsory Arbitration Schemes," Journal of Political Economy, 87,1 (1979) pp. 131-60.

Dertouzos, J.N. and J.H. Pencavel. "Wage Determinants Under Trade Unionism: The International Typographical Union," Stanford University (April 1980).

Diewert, W.E. "The Effects of Unionization on Wages and Employment: A General Equilibrium Analysis," Economic Inquiry, 12(2), pp. 319-39 (Sept. 1974)

Dunlop, John C. Wage Determination Under Trade Unions, MacMillan, New York and London (1944).

Epple, D., V.J. Hotz and A. Zelenitz. "Employment Contracts, Risk Sharing and the Role of Unions," Carnegie-Mellon Working Paper #70-79-80, (1980).

Farber, Henry. "Individual Preferences and Union Wage Determination: The Case of the United Mine Workers," Journal of Political Economy, 86,5 (1979) pp. 923-42.

__________. "Bargaining Theory, Wage Outcomes, and Occurrence of Strikes: An Econometric Analysis," American Economic Review, 68 (June 1978) pp. 262-71.

Freeman, Richard. "Individual Mobility and Union Voice in the Labor Market," American Economic Review (May 1976).

__________. "The Effect of Unionism on Worker Attachment to Firms," Journal of Labor Research, 1(1), pp. 29-61 (Spring 1980).

__________. "The Effect of Trade Unionism on Fringe Benefits," Naitonal Bureau of Economic Research Working Paper #292 (1978).

Hall, R.E. and D. Lilien. "Efficient Wage Bargains Under Uncertain Supply and Demand," American Economic Review 69(5) pp. 868-79 (Dec. 1979).

Hicks, J.R. The Theory of Wages, MacMillan: London (1966).

Johnson, H.G. and P. Miezkowski. "The Effects of Unionization on the Distribution of Income," Quarterly Journal of Economics (1970).

Lewis, H.G. Unionism and Relative Wages in the U.S., Chicago: University of Chicago Press (1963).

Marshall, A.P. Principles, Book V, Chapter VI.

Medoff, J. "Layoffs and Alternatives Under Trade Unions in United States Manufacturing," American Economic Review (June 1979).

Mincer, J. "The Economics of Wage Floors," American Economic Review, Papers and Proceedings of Sept. 1980 Meetings, 1981.

Pigou, A.C. Economics of Welfare, Book IV, Chapter V.

Rees, A. The Economics of Trade Unions, University of Chicago Press: Chicago (1962).

Reder, Melvin. "Job Scarcity and the Nature of Union Power," Industrial Labor Relations Review (1959), 13 pp. 349-62.

___________. "A Theory of Union Wage Policy," R.E. Stat. (1952), 34, pp. 34-45.

Rosen, S. "Trade Union Power, Threat Effects and the Extent of Organization," R.E. Stat. (1969).

______. "Unionism and Occupational Wage Structure in the U.S.," I.E.R. (June 1970).

______. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy 82 (Jan./Feb. 1974) pp. 34-55.

Ross, A.M. Theory of Union Wage Policy University of California Press: Berkeley (1948)

Sayles, Leonard and George Strauss. The Local Union, Harcourt, Brace and World: New York (1967).

Stevens, C. "On the Theory of Negotiation" Quarterly Journal of Economics 73, pp. 77-97 (1958).

Stiglitz, J. "Utilitarianism and Horizontal Equity: The Case for Random Taxation," mimeo (1980).

Weinstein, P.A. "The Featherbedding Problem," American Economic Review 54, pp. 145-52 (1964).