Carnegie-Rochester Conference Series on Public Policy 55 (2001) 229–264
North-Holland
www.elsevier.nl/locate/econbase

Deunionization, technical change and inequality*

Daron Acemoglu†
Massachusetts Institute of Technology
Philippe Aghion
Harvard University, Dept. of Economics, Cambridge, MA 02138
and
University College London
Giovanni L. Violante
University College London

Abstract

Over the past 25 years, the US and the UK experienced sharp increases in wage inequality and rapid deunionization. We argue that these two phenomena are related, and that skill-biased technical change has been an important factor in deunionization as well as in the rise in inequality. Skill-biased technical change causes deunionization because it increases the outside option of skilled workers, undermining the coalition among skilled and unskilled worker in support of unions. Our approach implies that although deunionization is not the underlying cause of the increase in inequality, it amplifies the direct effect of skill-biased technical change by removing the wage compression imposed by unions. We also show that deunionization may happen inefficiently.

1 Introduction

Over the past 25 years, the US and the UK experienced sharp increases in wage inequality and rapid deunionization. In the US, in 1980 the ratio of the 90th to the 10th percentile of the distribution of male weekly wages was

*Prepared for the Carnegie-Rochester Conference Series on Public Policy.

† *Correspondence to*: Philippe Aghion, Department of Economics, Littauer Center, Harvard University Cambridge, MA 02138 USA. Email: paghion@arrow.fas.harvard.edu

PII: S0167-2231(01)00058-6

2.7, and 24% of all private sector workers were unionized. By 1990, the 90-10 differential had risen to 3.5 and only 12% of private sector workers were unionized (see, e.g., Juhn, Murphy and Pierce, 1993, Farber and Krueger, 1992). In the UK, the ratio of the 90-10 wage differential was 2.4 and increased to 3.1 in 1990, while union density among male workers was 54% in 1980, and fell to 38% in 1990 (Gosling and Machin, 1995).

There is a variety of evidence that unions compress the structure of wages (see Reynolds, 1967, DiNardo, Fortin, and Lemieux, 1996, Card, 1996, and Fortin and Lemieux, 1997). Rees (1964, p. 64), for example, writes "...unions often adopt a policy of reducing wage differentials among occupations principally by bargaining for across-the-board wage increases...". Freeman and Medoff agree with this conclusion and write "...the data makes it clear that unions have a major impact on within plant inequality, and thus on the overall inequality" (1984, p. 82), and "...union standard rate policies tend to produce greater similarity in pay across establishments then does an unorganized labor market" (1984, p. 85).

Since unions compress the wage structure, many economists suspect that their decline may have been an important factor in the increase in inequality in the Anglo-Saxon economies (Freeman, 1991, DiNardo, Fortin, and Lemieux, 1996, Card, 1996, and Fortin and Lemieux, 1997). This conjecture receives casual support from the fact that the increase in inequality was sharper in non-union establishments (Gosling and Machin, 1995 for the UK, and Bratsberg and Ragan, 1997 for the US) and from evidence that wage inequality increased only little or not at all in many Continental European economies where unions still play an important role.[1]

Most economists, however, discount the role of unions in the increase in inequality for a variety of reasons. Figures 1, 2 and 3 plot measures of wage inequality and union density for the US, the UK and Canada (see the data appendix for sources and detailed definitions). In all three countries, wage inequality and union density move in tandem, but in all cases, deunionization appears to lag the increase in inequality. In the US, although unionization in the private sector continued its steady decline that started in the 1950s, overall unionization rates did not decline until the 1980s (see also Farber and Western, 2000, and Riddell, 1993, Table 4.1). In contrast, the rapid surge in inequality started during the early 1970s (see Juhn, Murphy and Pierce, 1993).[2] In the UK, wage inequality started its sharp increase in 1977,

[1]For example, Freeman (1988) reports that among OECD countries Denmark, Finland, Sweden and Belgium have experienced sustained increases in union density between 1970 and 1985, while Germany, France, Switzerland, Canada, Australia, Ireland, Norway, and Netherlands experienced stable union density. Among the same countries, wage inequality has increased sharply only in the UK and the US over this period, with moderate increases in Canada and Australia (OECD, 1993, table 5.2).

[2]This is the timing that emerges from the March Current Population Surveys and

while union density increased until 1979 (see also Gosling, 1998). Finally, in Canada the decline in unionization is not as pronounced and starts about five years after the increase in inequality.

Skill-biased technical change is the most popular explanation for the increase in inequality. Katz and Murphy (1992) argue that the increase in wage inequality can be explained by steady skill-biased technical change, interacting with the relatively slow growth in the supply of skills during the 1980s. A number of other studies emphasize the possible acceleration in the degree of skill bias of new technologies (e.g. Acemoglu, 1998; Aghion and Howitt, 1998; Autor, Katz and Krueger, 1998; Berman, Bound and Machin, 1998; Caselli, 1999; Galor and Moav, 2000, Krusell et al., 2000; Violante, 2000). Irrespective of whether there has been an acceleration in skill bias, there is widespread agreement that technical change over the past 30 years has been skill-biased (see for example Katz and Autor, 1999, Acemoglu, 2000).

In this paper, we advance the hypothesis that skill-biased technical change is at the root of deunionization as well as the rise in inequality. The basic argument we propose rests on three premises: 1) unions exist because they provide some benefits, either to the society as a whole, or simply to some group of workers, 2) wage compression across workers with different skills is a fundamental characteristic of unions, 3) there is skill-biased technical change. Our hypothesis is that when the degree of skill bias of new technologies is limited, the benefits provided by unions outweigh the costs of wage compression for skilled workers, who accept to work in unionized firms. Skill-biased technical change undermines the coalition between skilled and unskilled workers underlying unions because, by widening the productivity differentials, it increases the outside option—e.g., the competitive market return—of skilled workers and weakens their incentives to join the unionized sector.[3] Our approach therefore implies that although deunionization is not the primary cause of the surge in wage inequality, it *amplifies* the original effect of skill-biased technical change by removing the wage compression imposed by unions.

---

decennial Census data. DiNardo, Fortin and Lemieux (1996) report that in the May Current Population Survey the increase in wage inequality begins later, but Katz and Autor (1999) find than that wage inequality started increasing in the 1970s also using the May Current Population Surveys.

[3]Others economists also noted the tension between skilled and unskilled workers in unions. For example, Rees long ago wrote "...when the levelling policies of industrial unions outrun the underlying economic forces... important counterpressures are soon felt. The United Automobile Workers has had to negotiate special increases for skilled workers on more than one occasion... In a dramatic case of protest against wage levelling, the motormen in the NYC subway system temporarily broke away from their union, formed a separate craft union and conducted an effective strike for the widening of wage differentials." (1964, p. 64).

Unions are complex institutions and play a variety of different roles. Various authors emphasize different aspects of unions. The standard view is one in which unions are pure rent-seeking organizations (e.g., Oswald, 1985, and Lindbeck and Snower, 1988, especially pp. 82-83). They distort the socially optimal allocation of resources and represent an impediment to free contracting in the labor market. An alternative approach, initiated by Hirschman (1970), challenged this view: if unions are "bad institutions", why have they been so resilient? Why do they not compete among each other, like firms, with competition washing away their rents? One answer is that unions are institutions created as a response to particular forms of market failures or contract incompleteness. Freeman and Medoff (1984) and Freeman and Lazear (1995), for example, point out that unions improve employer-employee relationship and communication, and reduce labor turnover. Malcomson (1983) and Hogan (2001) emphasize unions' role in providing workers and firms a credible enforcement mechanism for state-contingent wage contracts.

We show that the theoretical validity of our hypothesis—that skill-biased technical change may cause deunionization—does not depend on the exact role that unions play. We develop our argument first with rent-seeking unions, where the main role of unions is to transfer rents from skilled to unskilled workers. We then generalize our results to the case where unions have an efficiency-enhancing role. To illustrate how efficiency-enhancing unions affect our results, we focus on the potential role of unions in encouraging training and providing insurance to workers. Although the positive implications of different models are similar, we find that when unions play an efficiency-enhancing role, deunionization may happen inefficiently: skilled workers cross-subsidize unskilled workers through their unionization decisions, and since skilled workers ignore this positive externality, they will tend to deunionize too soon. As a result, deunionization may lead to a deterioration in the allocation of resources in the economy.

An important aspect of our approach is the emphasis on the relationship between technology and workplace organization. When unions have an efficiency-enhancing role, they are useful organizations in conjunction with certain types of technologies. However, they become unsustainable when new technologies become more skill-biased. In emphasizing the interaction between technology and workplace organization, our approach is related to that of Piore and Sabel (1984), who argue that while unions improve the implementation of some technologies, they are fundamentally incompatible with others (see also Osterman, 1994). Piore and Sabel, however, do not emphasize skill-biased technical change and do not develop their argument formally.

Two empirical facts appear important for a theory of deunionization. Using data from the Panel Study of Income Dynamics (PSID), we compute

union membership rates in the US for male workers between 1976 and 1992. In Figures 4 and 5, we plot union membership rates for three age groups, 18-29, 30-45, and > 45 years old relative to the average unionization rate in the economy (Figure 5 refers only to manufacturing). Until the late 1980s, there has been a much more rapid deunionization among young workers, while Figure 5 displays a slight decline also for middle-aged workers after 1988. A related fact is reported in Figure 6 where we plot college enrollment rates and union membership rates in the US. This figure points to a negative relationship between college enrollments and unionization. As union density begins its rapid decline in the years 1979-1980, college enrollments start to increase. This finding fits well with the age pattern of deunionization described above: new entrants in the labor force are acquiring more formal schooling which leads them towards jobs and/or sectors where the incentives to organize unions are weak.

We think that these two facts call for a modelling strategy incorporating two distinct types of deunionization: a reduction in the inflow into unions by young workers who enroll in college and work in non-unionized jobs—what we refer to as *ex ante deunionization*—, and a later increase in the outflow of older workers from unions—*ex post deunionization*. Our model incorporates both types of deunionization.

Finally, it is important to stress that we do not mean to argue that technical change was the only factor responsible for deunionization. There are at least three other important candidates. The most popular explanation for deunionization is the changing legal environment and increasing management opposition. It is clear that the anti-union efforts of Thatcher in the UK and Reagan's reaction to the strike by the Professional Air Traffic Controllers Organization (PATCO) in the US were instrumental in reducing unions' influence. Nevertheless, a careful look at the timing of the events show that the key forces undermining unions were already under way, when the legal framework began to turn unfavorably against unions. Figure 2 shows that deunionization in the UK started probably before the first Thatcher government was elected. For the US, Farber and Western (2000) document a decline in union organizing activity already before the PATCO strike and cast some doubt on the common view that this event precipitated the falling trend in unions' organizing activity.

The second major explanation for deunionization is that, because of structural reasons, organizing union activity is cheaper in some industries than in others, and there have been major changes in the industrial composition of production towards those industries (e.g. services) where organization is more expensive. However, both Pencavel (2000) for the UK and Farber and Krueger (1992) for the US conclude that industrial demand shifts can account only for a fraction of the total decline in union density. Most of the

sectors of traditional union strength had already contracted substantially before the wave of sharp deunionization started. Finally, deunionization may have resulted from increased competition among firms, reducing rents that could be appropriated by unions. We develop and empirically investigate this explanation in our companion paper, Acemoglu, Aghion, Machin and Violante (2001).[4]

The rest of the paper is organized as follows. Section 2 describes the economic environment; Section 3 presents a model of rent-seeking unions; Section 4 presents two versions of the efficiency-enhancing view of unions, one based on insurance and another on training. Section 5 concludes the paper. In the Appendix, we report the data sources for the figures.

2 The environment

The economy is populated by a continuum of workers of measure 1. Workers are risk-neutral and consume their wage income w net of the cost of skill acquisition. In Section 4.2, we extend this model to allow for a strictly concave utility function, $u(\cdot)$.

There are two sectors in the economy, A and B, that produce the same good with different technologies. Both technologies are linear in efficiency units of labor, but differ in productivity and skill requirements. Workers can be "skilled" or "unskilled". We assume that skilled workers produce $y_s^A = A\eta$ and $y_s^B = \eta$ in the two sectors, where $A > 1$. Unskilled workers cannot operate sector A technology, so their productivity is $y_u^A = 0$ in sector A, and $y_u^B = \alpha$ in sector B, with $\eta > \alpha > 0$. We also assume that although the production technologies are linear, each firm needs to employ at least a continuum of workers of mass $\varepsilon > 0$. The role of this assumption will become clear below.

To become skilled, workers need two skill components: a "specific" component and a "general" component. The specific component is necessary to produce in each sector, and it is firm-specific. We assume that each worker has to incur a cost $\overline{e} > 0$ to acquire this firm-specific skill component in either sector. Specific skills can be acquired at any point in time.

The general skill component, which is necessary for production in sector A, can be innate, or acquired by education before workers make their sector choice. A worker who does not invest in education possesses this general ability with probability $\phi < 1/2$. Those who invest in education acquire

[4]Two other explanations for the decline in unions that need to be mentioned are (1) the hypothesis that the welfare state or private companies are now providing the services traditionally provided by unions (Neumann and Rissmann, 1984, Kochan, Katz and McKersie, 1986, and more recently van Leewen, 1997), and (2) changing public opinion towards unions (Lipset, 1986).

this general skill with probability 1. Finally, we assume that workers have different costs of education, κ, and denote the distribution of this cost in the population by $G(\kappa)$.

To motivate these assumptions, we can think of the two sectors as the software industry and a traditional manufacturing industry (e.g. textile). To work in the former, one needs a basic knowledge of how to operate computers, while to work in the latter one needs to know how to operate specific textile equipments. However, the tasks associated to jobs in the software industry require some key cognitive, analytical and numerical skills, which correspond to the general component in our model. Such skills are rewarded, but they are not essential in the traditional textile sector.

We make two informational assumptions. First, while worker productivity is common knowledge, a worker's education cost κ and skill level are not directly observed by any other agent in the economy. Thus, wage contracts cannot be made contingent on skills or schooling costs, but only upon productivity. Second, workers who have not acquired formal education discover their skill level only *after* they have decided between entering sector A or sector B. This will allow us to distinguish between ex ante and ex post de-unionization, a potentially important feature of the data, as argued in the Introduction.

In both sectors, firms compete by offering wage contracts of the form

$$w^i\left(y^i\right) = \gamma + \beta y^i, \tag{1}$$

where i = A, B, $\beta \geq 0$ and $\gamma \geq 0$.[5] The technology assumptions ensure that only skilled workers will be employed in sector A. Since the production technology in this sector is linear, all workers will be paid their marginal product—i.e., $w_s^A = A\eta$. In sector B, there will be both skilled and unskilled workers. The linear technology implies that the group of employees have to be paid their collective marginal product (otherwise the firm would go bankrupt).[6] However, because switching to the other sector is costly, it is possible for some workers to be paid more than their marginal product, while others receive less—i.e., cross-subsidization can arise. This feature allows unionization in sector B. A union makes a wage contract offer to the firm of the form (1), and the firm can either accept it (in which case the union contract overrides the previous wage contract offered by the firm) or exit and make zero profits.

In what follows, we distinguish between two different types of unions that can be formed in sector B.

[5]Given the two point skill distribution, the assumption of a linear contract is not restrictive.

[6]The analysis becomes more involved when there are decreasing returns to scale. This is the case we analyze in our companion paper.

1. **Rent-Seeking Unions:** a coalition of workers choose a wage contract and impose it upon the firm. The wage schedule is determined by pure majority voting among all union members. Taking advantage of the ex post skill heterogeneity and the switching costs, each type of worker tries to extract rents from the other type. This kind of union has no efficiency benefits.

2. **Efficiency-Enhancing Unions:** a coalition of workers form a union that also solves a market failure problem. We consider two different channels whereby unions can improve efficiency: unions inducing training and unions providing insurance. In the first case, we suppose that a training technology is available in sector B to turn unskilled workers into skilled ones. When training is non-contractible, a union can be essential for such a technology to be used in equilibrium (see Acemoglu and Pischke (1999a, 1999b)). Alternatively, we discuss the case where workers are risk-averse. Given that each worker in sector B faces uncertainty over her skill type, when insurance markets are incomplete, equilibrium wage contracts can be designed to offer at least some degree of insurance, and unions are a vehicle to provide such insurance.

The two cases of rent-seeking and efficiency-enhancing unions span a wide range of union behavior and union practices. In the rent-seeking case, unions are coalitions of workers that can negotiate wages with firms. In particular, we assume that unions make a take-it-or-leave-it offer to firms as in the standard right-to-manage model of unions (see for example Booth, 1995). In the efficiency-enhancing case, unions are institutional arrangements to enforce wage contracts that involve some degree of wage compression between skilled and unskilled workers, in order to provide training or insurance. Absent any enforcement mechanism, ex post firms would try to escape the effects of such wage contracts in order to increase their profits: they could try to poach skilled workers from other firms by offering a wage above the union wage (which is profitable since union wages are lower than the marginal product for skilled workers), or they could fire unskilled workers who are paid more than their marginal productivity. Unions are institutional arrangements aimed at avoiding such deviations and thereby sustaining the "good" equilibrium with training or insurance:[7] in particular unions can prevent firms from firing unskilled workers by establishing binding *firing restrictions*, and can prevent firms from hiring skilled workers from other firms by imposing

[7]We do not mean to imply that there are no other institutional arrangements that would help with the same market imperfections or contract incompleteness. Our claim is the more modest one, that unions, which exist for a variety of reasons, can also play these useful roles since they already regulate hiring and firing in practice.

binding *ports-of-entry hiring policies.*[8]

Before proceeding to the analysis, it is useful to outline the exact timing of events:

- Workers find out their education cost κ and choose whether to obtain education.
- Non-unionized firms make wage contract offers of the form (1) as a function of worker productivity.
- Workers join firms in sector A or B, and incur the firm-specific cost $\overline{e}$.
- Workers decide whether to unionize or not.
- Those who have not obtained education find out whether they are skilled or not.
- If there is a union, unionized workers vote over the wage policy, i.e., over γ and β in the wage contract. Then the firm decides whether to accept the contract offer or not. If it accepts this offer, it is committed to pay the contracted wage to all workers who stay.
- At this point, workers can also decide to switch firms or sectors and if they do so, they incur the firm-specific cost $\overline{e}$ again.
- Production and consumption take place.

At this point, we can also briefly characterize the equilibrium without unions for future reference. All workers will be paid their marginal product, so we have $w_s^A = A\eta$, $w_s^B = \eta$ and $w_u^B = \alpha$. Skilled workers in sector B will stay there as long as $(A-1)\eta \geq \overline{e}$.

3 Rent-seeking unions

Let us first derive the expected net values of a worker entering sectors A and B respectively. First, note that in sector A, there are only skilled workers. Recall that firms have to make zero profits, so all workers have to be paid their marginal product. Consequently, there will not be unions in this sector and $w_s^A = A\eta$. On the other hand, there could be unions in sector B. We will look for an equilibrium with unionization in this sector, so we write wages

[8]Empirically, besides wage bargaining, hiring and firing restrictions are among the most traditional roles of unions. Millward et al. (1992), based on the 1990 *Workplace Industrial Relation Survey* for the UK (Tables 7.16 and 7.17), report that between 40% and 50% of unions negotiate over the size of redundancy pay and between 30% and 40% of unions negotiate over recruitment practices.

as $\gamma + \beta y^i$, where y^i is the productivity of the worker in question. We adopt the convention that if the only wages the union can demand are equal to the competitive wages, there will be no unionization (for example, due to the existence of a positive cost of unionization).

The values for educated workers from entering into sectors A and B respectively, are:

$$\begin{aligned} E^A &= A\eta - \bar{e} \\ E^B &= \gamma + \beta\eta - \bar{e}. \end{aligned} \tag{2}$$

Next, the corresponding values for uneducated workers, V^A and V^B, are:

$$\begin{aligned} V^A &= -\bar{e} + (1-\phi)\max\{\gamma + \beta\alpha - \bar{e}, 0\} + \phi\max\{A\eta, \gamma + \beta\eta - \bar{e}\}, \\ V^B &= -\bar{e} + (1-\phi)(\gamma + \beta\alpha) + \phi\max\{A\eta - \bar{e}, \gamma + \beta\eta\}. \end{aligned} \tag{3}$$

These expressions take account of the fact that an uneducated worker will be revealed to be skilled or unskilled, at which point he can decide to switch sectors. The max operators incorporate this choice. For example, the second line incorporates the fact that a worker is revealed to be skilled with probability ϕ, at which point he can either stay in sector B and receive the equilibrium contractual payoff, (1), $\gamma + \beta\eta$, or switch to sector A and receive his full marginal product $A\eta$, but also incur the switching cost $-\bar{e}$. In writing these expressions, we implicitly used the fact that there will not be any worker who switches firms within the sector (since this will be costly, but without any benefits), and that all firms in sector B offer the same contract (otherwise, all workers would go to the firm offering the most attractive contract).

Majority voting among union members favors unskilled workers (recall that $\phi < 1/2$), so we have: $\gamma + \beta\eta < A\eta$, and $E^A > E^B$. Hence, all educated workers will choose sector A. Similarly, the fact that unions in sector B favor unskilled workers implies that: $w_u^B = \gamma + \beta\alpha > \alpha$. Then the condition: $\alpha > \bar{e}$, is sufficient to ensure that a worker revealed to be unskilled in sector A prefers to switch to sector B and be employed. We shall assume this condition throughout the remaining part of our analysis. Now, comparing between V^A and V^B, one can show that ex ante uneducated workers prefer to enter sector B. To see this note that under the assumption that $\alpha > \bar{e}$, we have: $V^A - V^B = (1-\phi)\bar{e} - \phi[A\eta - \max\{A\eta - \bar{e}, \gamma + \beta\eta\}] < 0$, since the term in square brackets cannot exceed $\bar{e}$, and $\phi < 1/2$. Therefore, all educated workers will enter sector A, while all uneducated workers will choose sector B. Finally, the comparison between $E^A - \kappa$ and V^B will determine whether a worker chooses ex ante whether or not to acquire education. To characterize the full equilibrium, we now have to determine γ, β and whether skilled workers will choose to quit sector B and switch to sector A.

Recall that workers choose employers before they know their skill level, and since all firms have to employ at least a continuum of workers of mass ε, by the law of large numbers all firms in sector B will have a fraction ϕ of their employees skilled and the rest unskilled.[9] The median union member will therefore be unskilled and use his voting power to extract rents from skilled workers. The problem of the unskilled median voter is therefore:

$$\max_{\gamma,\beta} \{\gamma + \beta\alpha\} \tag{4}$$

$$s.t.$$

$$\gamma + \beta\eta \geq A\eta - \bar{e}, \tag{5}$$

$$\pi = -\gamma + [1 - \beta]\, Ey^B \geq 0, \tag{6}$$

where π is the firm's profit and Ey^B is the average productivity in the firm. The first constraint is the *no-quitting* condition for skilled workers and it requires that the net income of a skilled worker from staying in sector B (in the union) be greater than the utility of moving to sector A net of the switching cost $\bar{e}$. This condition must hold, since otherwise all skilled workers would leave sector B, which in turn would hurt unskilled workers and leave them with a wage that does not exceed α. Also note that workers in sector B will also incur the same switching costs if they move to another firm in sector B, so that this latter opportunity is dominated by moving to sector A. The second constraint simply states that the firm must make non-negative profits. If not, the firm would simply shut down.

When they expect all workers to stay, firms in sector B will evaluate their average productivity as $Ey^B = \phi\eta + (1 - \phi)\,\alpha$. Since the equilibrium contract is one that satisfies the no quitting constraint for skilled workers, (5), in equilibrium neither skilled nor unskilled workers will leave the firm. Then, the zero profit constraint, (6), which binds in equilibrium, implies:

$$\gamma = (1 - \beta)\,[\phi\eta + (1 - \phi)\,\alpha]\,. \tag{7}$$

Note that the competitive wage structure corresponds to the extreme case where $\gamma = 0$ and $\beta = 1$, while if $\beta < 1$ and $\gamma > 0$, the contract entails wage compression, with the skilled workers with productivity η cross-subsidizing unskilled workers with productivity α. The case $\beta = 0$ corresponds to a flat wage-skill profile.[10] The union will never choose $\beta > 1$, since this would imply that $w^B(\alpha) < \alpha$, thereby making it profitable for unskilled workers to revert

[9] In the absence of the assumption that firms have to hire a continuum of workers, there could be one firm-one-worker relationships. Our assumption rules this possibility out.

[10] We are implicitly allowing $\beta < 0$. This simply prevents us from carrying a 'max' operator around. For example, with the restriction that $\beta \geq 0$, in Proposition 1, we would have $\beta^* = \max\left\{1 - \frac{\bar{e}-(A-1)\eta}{(1-\phi)(\eta-\alpha)}, 0\right\}$.

to the competitive wage structure by leaving the union. The equilibrium for this economy is described in

Proposition 1. *With rent-seeking unions, there is a unique equilibrium where, in sector A, $w^A(\eta) = A\eta$. In sector B, there exists a threshold value*

$$\eta^* \equiv \frac{\overline{e}}{A-1}$$

such that for $\eta > \eta^*$ firms are not unionized and pay $w^B(\eta) = \eta$, $w^B(\alpha) = \alpha$, while for $\eta \leq \eta^*$ firms in sector B are unionized and the union imposes the wage contract (1) with $\beta^* = 1 - \frac{\overline{e}-(A-1)\eta}{(1-\phi)(\eta-\alpha)} \leq 1$, and $\gamma^* = (1-\beta^*)[\phi\eta + (1-\phi)\alpha]$. All workers with cost of schooling less than κ^* acquire education and enter sector A, while the rest of the workers enter sector B and join unions, where

$$\kappa^* \equiv A\eta - [\phi\eta + (1-\phi)\alpha]. \tag{8}$$

No worker quits sector A after entry. For $\eta \leq \eta^*$, no worker quits sector B, while when $\eta > \eta^*$, skilled workers switch to sector A ex post.

Proof. That the above allocation is an equilibrium is straightforward. We have already shown that $V^B > V^A$ and $E^A > E^B$, so that all uneducated workers enter sector B and all educated workers go to sector A. In sector B, the median voter is an unskilled worker since $\phi < 1/2$, so he will solve the maximization problem (4), which has a solution with both constraints binding. This gives the expressions for β^* and γ^*. If β^* is greater than 1, skilled workers cannot be convinced to stay in sector B, so there will be no unionization. This happens when $\eta > \eta^*$. Finally, comparison of E^A and V^B yields the cut-off value κ^*. To show uniqueness, note that as long as (5) holds, skilled workers in sector B will never leave their firms. Therefore, firms will always accept the contract offer (1). This implies that any allocation that is not a solution to (4) cannot be maximizing the utility of an unskilled worker, and hence cannot be part of an equilibrium. This establishes uniqueness.□

The key result of this Proposition is that unions are likely to arise in equilibrium for low values of the productivity of skilled workers (η) and the productivity of technology in the education-intensive sector (A). Unskilled workers can induce skilled workers in sector B to accept a wage below their marginal product because of the positive switching cost. However, for this wage agreement to be sustainable in equilibrium, the outside option of sector B skilled workers, $A\eta$, cannot be too large, otherwise the only way to convince skilled workers to stay in sector B would be for unskilled workers to be paid below their marginal product. Finally, once skilled workers leave sector B,

there is no point for unskilled workers to unionize in this economy,[11] and we refer to this case as "deunionization".[12]

3.1 *Discussion*

3.1.1 *Technical change, deunionization and inequality*

The degree of skill bias in technology in this economy can be measured in two ways. First, by the productivity in the education-intensive sector A; second by the difference between η and α. The effect of a rise in A is similar to a rise in η with α constant. Both comparative statics lead to two types of deunionization. First, there is *ex ante deunionization*: more workers acquire schooling and enter sector A, which has become relatively more attractive. This can be seen from (8) which shows that κ^* is increasing in A and η. With ex ante deunionization, unions continue to exist, but their size shrinks.

Second, there is *ex post deunionization* leading to the collapse of unions. The equilibrium wage contract in Proposition 1 implies that β^* is increasing in A. The slope of the wage contract β^* is also increasing in η for $\overline{e} > (A-1)\alpha$, which is always verified as long as unions exist, i.e., as long as $\eta < \eta^*$. Thus, as the relative productivity of skilled workers or of the education-intensive sector increases, it becomes more difficult to maintain the wage compression required for rent-extraction, and when β^* eventually exceeds 1, unions become unsustainable, and all skilled workers leave the union.

It is also useful to discuss briefly the implications of skill-biased technical change on inequality through deunionization, and contrast this with the effect of such technical change on inequality in an economy without unions. Recall from above that in the competitive economy, there are three wages in equilibrium, $A\eta$ for workers in sector A, η for skilled workers in sector B, and α for unskilled workers in sector B. Therefore, an increase in A or η increases inequality by raising the highest wage in the distribution, $A\eta$, without affecting the lowest wage rate, α, paid in sector B. When unions are present, we have instead that

$$w^B(\alpha) = \alpha + \frac{\phi}{1-\phi}\left[\overline{e} - (A-1)\eta\right]. \tag{9}$$

The total rents extracted by the union from skilled workers are equal to $\phi\left[\overline{e} - (A-1)\eta\right]$, and these rents are redistributed equally among the unskilled, so that the second term of (9) is the union premium in excess of the

[11] It is straightforward to generalize the model to include a cost of unionization, in which case unskilled workers would strictly prefer not to unionize.

[12] We choose this terminology since we are using this static economy to think of the various stages of a more dynamic economy, which starts with low η, and is unionized, and becomes "deunionized" as η increases beyond the critical threshold.

marginal productivity of unskilled labor. From (9), it follows immediately that skill-biased technical change—an increase in A or η— induces a fall in the wages at the bottom of the wage distribution. The effect of technological change on inequality is therefore *amplified* by changes in unionization. Moreover, the real wage of the unskilled falls, a prediction that fits well with the observed changes in the wage structure in the US (see Katz and Autor (1999) for a survey).

3.1.2 *Welfare*

Consider the problem of a social planner who faces the same informational constraints as a decentralized economy. This planner will allocate workers between sectors before the general skill component of each individual is realized. She makes three choices: first, the threshold level of learning cost κ^{SP} below which workers are allocated to the education-intensive sector; second, the ex ante sectoral choices for workers; third, whether to move ex post all the skilled workers from sector B to sector A.

It is immediate to see that $\kappa^{SP} = \kappa^{\star}$ and that ex post reallocation of skilled workers (from sector B to sector A) will take place if $\eta > \eta^{SP} = \eta^{\star}$. Suppose indeed that we are in the case $\eta \leq \eta^{SP}$, then it is easy to see that $V^A < V^B$. Now, for $\eta > \eta^{SP}$ and as long as $\phi < 1/2$, unskilled workers will enter sector B ex ante, and those who turn out to be skilled ex post will be optimally reallocated to sector A—as was already the case in the decentralized economy with unions.

The equilibrium with unions is therefore socially efficient. This result is intuitive: because of the zero profit condition, the rent-seeking union does not change the total labor share expected by workers upon entry in sector B, it just redistributes this share ex post from skilled to unskilled workers. Under risk-neutrality, the redistribution has no distortionary effect. Hence, despite rent seeking by unskilled workers, the equilibrium allocation is socially optimal. Unionization reduces income inequality in the economy, compared to the decentralized economy without unions, but does not distort the allocation of resources.

4 Efficiency-enhancing unions

We now analyze an economy in which unions have an efficiency-enhancing role. To contrast this type of union with the rent-seeking unions analyzed in the previous section, we now assume that contract offers made by firms are binding, so unskilled union members in sector B do not directly negotiate wages with firms. Unions' role is simply to monitor recruitment and firing by sector B firms.

4.1 *Training*

Suppose that employers in sector B can train unskilled workers. A fraction ϕ of workers who are hired are already skilled and do not need to be trained. The remaining workers can be trained at per capita cost τ and a fraction $\lambda \in (0,1)$ of those become skilled. We impose $\tau < \lambda(\eta - \alpha)$ so that training is socially beneficial.

We assume that training decisions are *non-verifiable ex post.* Therefore, the standard "Becker solution" whereby workers fully pay for general training is not implementable: firms in sector B would promise training ex ante in return for a payment,[13] but then ex post they would have a profitable deviation by not honoring their promise and not training them. This implies that there is no equilibrium with worker-sponsored training. Nevertheless, firms may be willing to invest in the training of their employees. They would do so only when the wage structure is sufficiently compressed, so that they make a profit from "transforming" unskilled workers into skilled workers (see Acemoglu and Pischke (1999a,b)). This will also imply that employers can increase profits by poaching skilled workers from other firms, and by laying off unskilled workers who are paid above their marginal product. The role of unions in this economy will be to regulate hiring and firing to prevent such deviations. In the absence of unions, the wage structure would be given by workers' marginal products in both sectors, and as a result, there would be no training investments.

The timing of events now is:

- Workers find out their education cost κ and choose whether to obtain education.
- All firms make wage contract offers of the form (1) as a function of worker productivity. Whether there will be training or not is not part of this contract, since training is non-contractible.
- Workers join firms in sector A or B, and incur the firm-specific cost $\bar{e}$.
- Those who have not obtained education find out whether they are skilled or not, and the firm decides whether to train those who are unskilled.
- Trained workers find out whether they have acquired the necessary skills.

[13] Alternatively, if workers were productive during the training period, firms would pay less than workers productivity to cover the training cost.

- Workers decide whether to switch firms or sectors and if they do so, they incur the firm-specific cost $\overline{e}$ again. Firms hiring and firing decisions at this point are regulated by unions.

- Production and consumption take place.

Notice the difference between the timing of events here and that with rent-seeking unions. There is no voting over the wage contracts now, and the only role of unions is to regulate hiring and firing. This implies that wages will be determined to attract workers, hence sector B firms will maximize the expected value of workers that they want to attract, and this will force them down to zero profits. This implies that the equilibrium in sector B can be characterized as the solution to the maximization problem

$$\max_{\gamma,\beta} \{\gamma + \beta [\phi\eta + (1-\phi)\alpha]\}, \tag{10}$$

subject to three constraints: zero profits for the firms, a training constraint, and no quitting constraints for skilled workers, which is the same as in the previous section, given by (5).

The training constraint for the firm requires that the firm increases its profits by training. Hence

$$\lambda(1-\beta)(\eta-\alpha) \geq \tau. \tag{11}$$

This constraint states that for the firm to offer training to an unskilled worker, there needs to be a certain amount of wage compression. More explicitly, the expected increase in productivity for a worker net of the increase in the wage rate, $\lambda\left[\left(\eta - w^B(\eta)\right) - \left(\alpha - w^B(\alpha)\right)\right]$, has to exceed the training costs τ. When the wage structure is at the competitive level, i.e. $w^B(\eta) = \eta$ and $w^B(\alpha) = \alpha$, this constraint is violated. So, wage compression, i.e., $\beta < 1$, is necessary for training.

The zero profit condition for firms when training is provided is

$$[\phi + (1-\phi)\lambda](1-\beta)\eta + (1-\lambda)(1-\phi)(1-\beta)\alpha - \gamma - (1-\phi)\tau \geq 0. \tag{12}$$

This expression is intuitive: the firm incurs the training cost τ for a fraction $(1-\phi)$ of workers and pays the flat component of the wage contract γ to all workers. Following training, a fraction $[\phi + (1-\phi)\lambda]$ of workers are skilled, and the remaining fraction of workers are unskilled.

The solution to the maximization problem will depend on which constraints bind. It is straightforward to see that as long as all three constraint hold, any transfer from skilled to unskilled workers, or vice versa, will leave the ex ante utility of workers, as given by (10), unchanged. Competition among firms will always lead to exactly zero profits in equilibrium, i.e. (12)

holds with equality. So without loss of any generality, we focus on the case where it is the no quitting constraint, (5), to hold as an equality. This gives the equilibrium wage contract as

$$\begin{aligned} \beta^* &= 1 - \frac{\overline{e} - (1-\phi)\tau - (A-1)\eta}{(1-\lambda)(1-\phi)(\eta-\alpha)} \text{ and} \\ \gamma^* &= (1-\beta^*)[\eta - (1-\lambda)(1-\phi)(\eta-\alpha)] - (1-\phi)\tau. \end{aligned} \tag{13}$$

However, this contract will be offered in equilibrium only if it also satisfies the training constraint, (11). Define

$$\overline{\beta} \equiv 1 - \frac{\tau}{\lambda(\eta-\alpha)}$$

For training to be incentive compatible, there needs to be enough wage compression, in particular, we need $\beta^* \leq \overline{\beta}$. Then, we have

Proposition 2. *There is a unique equilibrium allocation, where sector A wages are given by $w^A(\eta) = A\eta$. In sector B, the equilibrium wage structure is characterized by the cutoff level $\eta^* \equiv \frac{\overline{e} - \frac{(1-\phi)}{\lambda}\tau}{A-1}$; such that: (a) for $\eta > \eta^*$ there is no unionization and firms pay $w^B(\eta) = \eta$, $w^B(\alpha) = \alpha$; (b) if $\eta < \eta^*$, there is unionization and training. All workers with cost of schooling less than κ^* acquire education and enter sector A, while the rest of the workers enter sector B and join unions, where*

$$\kappa^* = \begin{cases} \kappa_1^* = A\eta - [\phi\eta + (1-\phi)\alpha + (1-\phi)(\lambda(\eta-\alpha) - \tau)] & \text{if} \eta \leq \eta^* \\ \kappa_2^* = A\eta - [\phi\eta + (1-\phi)\alpha] & \text{if} \eta > \eta^* \end{cases}.$$

No worker quits sector A after entry. For $\eta \leq \eta^*$, no worker quits sector B, and when $\eta > \eta^*$, skilled workers switch to sector A ex post.

Proof. The proof is similar to that of Proposition 1. Maximization of (10) has a solution where both the zero profit constraint, (12) and the no quitting constraint, (5) are binding. Hence the claim that there is a unique equilibrium allocation—not a unique equilibrium wage contract. This solution corresponds to (13). However, this is a solution only when it also satisfies the training constraint, (11), which requires that $\beta^* \leq \overline{\beta}$. Solving for the value of η that satisfies the equation $\beta^* = \overline{\beta}$, yields the cut-off value η^*; when $\eta \geq \eta^*$, (11) cannot be satisfied and therefore there is no unions in this case. The values for κ^* follow from comparing E^A and V^B respectively when $\eta < \eta^*$ (training) and when $\eta \geq \eta^*$ (no training). Uniqueness follows again since any allocation that is not a solution to this maximization problem cannot be an equilibrium.□

It is useful to reiterate the role of unions in this economy. Sector B firms are making zero profits, and $\beta < 1$. This implies that firms are paying above marginal product to unskilled workers. They can therefore increase profits by firing unskilled workers. But the anticipation of such firing would make workers unwilling to enter into this contract, and there would be no training in equilibrium. Unions, by preventing firms from firing unskilled workers and poaching skilled workers from other firms, support the training equilibrium. Although there may be other institutional arrangements that could also support such an allocation, given the frequent involvement of unions in hiring and firing, this seems a natural role for unions to play in practice. It is also interesting to note that the equilibrium allocation would have been identical if unskilled workers voted over the wage contract as in the previous section. This is the reason why we emphasized that unions are not involved in wage negotiations in this case—there are other reasons for wage compression, which lead to identical results.

An important result from the above analysis is that skill-biased technological change again makes unionization harder to maintain—this time because the wage compression necessary for training becomes incompatible with keeping skilled workers in sector B. Interestingly, Osterman reaches a similar conclusion to our analysis and relates deunionization to technical change; he writes "...the combination of technical change and the increased educational level of the workforce may alter firms' calculation of the best locus for training and undermine the traditional reliance on job ladders and closed internal markets" (1994, p. 302).

4.1.1 *Skill-biased technical change*

The effects of skill-biased technical change on wage inequality, unionization, and education, are similar to before. Consider an increase in A, or in η with α constant. The first effect of these changes is to increase inequality, as in the competitive economy, but also cause ex ante deunionization—fewer workers join sector B.

In addition, skill-biased technical change can cause ex post deunionization and transform the wage structure more radically. To see this, note that in the economy without unions, increases in A or η would lead to an increase in inequality, but no change in the bottom part of the wage distribution, exactly as before. In the economy with unions, this technical change will eventually lead to the violation of the no-quitting constraint— η would eventually exceed η^*. As unions collapse, wages at the bottom of the distribution would fall from $\overline{\gamma} + \overline{\beta}\alpha$ to α. This is an interesting contrast to the model with rent-seeking unions. There, with deunionization, there was no discrete jump in the wages of unskilled workers, because deunionization happened when

β^* reached 1. Here, deunionization happens before β^* reaches 1 because a certain degree of wage compression—i.e., $\beta \leq \overline{\beta}$—is necessary to support training and hence unions. This leads to a discrete decline in the wages of unskilled workers in sector B.

Another interesting consequence of deunionization is the amplification of the effect of skill-biased technical change on education acquisition. When the level of η rises but remains below η^*, the threshold for education, $\kappa_1^*(\eta)$, rises smoothly. This in turn induces ex ante deunionization, driven by the higher return to education. As η rises beyond η^*, unions disappear, the net return to education jumps discretely, and so does the equilibrium number of educated workers. Hence, deunionization amplifies the impact of skill-biased technical change on formal schooling, while at the same time reducing union-supported on-the-job training.

4.1.2 *Welfare*

The above discussion also has implications for the welfare comparison between the unionized and the competitive economy. Consider the problem of a social planner who faces the same informational constraint as the decentralized economy, i.e., the planner must allocate workers between sectors before unveiling the realization of the general skill component for each individual. In sector B, the planner can apply the training technology to all the unskilled workers at per-capita cost τ. The planner will allocate workers between education and training by choosing: first, the threshold level of learning cost κ^{SP} below which workers are allocated to the education-intensive sector, and second, the threshold level η^{SP} above which all the trained workers in sector B should be moved to sector A.

It is immediate to see that the solution of the planner's problem requires $\kappa^{SP} = \kappa_1^* < \kappa_2^*$ and it implies that ex post mobility of skilled workers will take place if $\eta > \eta^{SP} \equiv \frac{\bar{e}}{A-1} > \eta^*$. This result means that as long as unions exist, i.e. $\eta < \eta^*$, the equilibrium allocations are Pareto optimal, however there will exist values of η such that the decentralized economy is not unionized, and hence does not train workers, while the social planner prefers training. This implies that in the decentralized economy deunionization (and hence the end of training) occurs inefficiently. The competitive economy without unions does not make use of the training technology, thus it has an inefficiently high number of educated workers.

More formally, the fact that deunionization happens too soon in this economy implies that over a certain range of parameter values a small increase in the productivity of skilled workers, η, will lead to a decline in aggregate output. To see this, let Y^U and Y^C denote aggregate output in the unionized

and competitive economy respectively. Then:

$$Y^U = -\overline{e} - \int_0^{\kappa_1^*} \kappa dG(\kappa) + G(\kappa_1^*) A\eta + [1 - G(\kappa_1^*)] \\ [\phi\eta + (1-\phi)\alpha + (1-\phi)(\lambda(\eta - \alpha) - \tau)],$$

$$Y^C = -\overline{e} - \int_0^{\kappa_2^*} \kappa dG(\kappa) + G(\kappa_2^*) A\eta + [1 - G(\kappa_2^*)] [\phi\eta + (1-\phi)\alpha].$$

Using the equilibrium values for κ_1^* and κ_2^*, solving the integrals by parts, and evaluating output exactly at the cut-off point η^*, we obtain:

$$Y^U - Y^C = \kappa_2^* - \kappa_1^* - \int_{\kappa_1^*}^{\kappa_2^*} G(\kappa) d\kappa > 0,$$

where the last inequality follows from the fact that $G(\kappa) < 1$. So as η increases beyond η^*, aggregate output falls by a discrete amount. This is because unions are now playing a useful role in supporting training investments, and once unions disappear, firm-sponsored training collapses.

4.2 *Insurance*

Suppose workers are risk-averse, with a strictly concave utility function $u(\cdot)$ defined over income net of the costs of skill acquisition; in other words, we assume that $\overline{e}$ and κ are monetary costs. As before, workers entering sector A face uncertainty over their skill level. Because they are risk-averse, they would benefit from insurance against that risk. The most straightforward solution may be outside insurance, but in practice there are a number of reasons for why such insurance may be impossible. For example, workers and firms may be able to collude and hide the productivity of workers, receiving payments from insurance agencies.

We consider the extreme case where the only type of insurance available to workers is within the firm-worker relationship, through a wage contract that induces wage compression, with skilled workers being paid less than their marginal product, while unskilled workers receive more. However, firms will not provide this type of insurance by themselves: to the extent that wage compression entails a loss for the firm on each unskilled worker it employs, ex post the firm would find it profitable to deviate by firing unskilled workers and produce only with skilled ones. We will see below that unions can support firm-provided insurance by regulating hiring and firing.

The timing of events is now:

- Workers find out their education cost κ and choose whether to obtain education.
- All firms make wage contract offers of the form (1) as a function of worker productivity.

- Workers join firms in sector A or B, and incur the firm-specific cost $\bar{e}$.
- Those who have not obtained education find out whether they are skilled or not, and decide whether to switch firms or sectors and if they do so, they incur the firm-specific cost $\bar{e}$ again. Firms hiring and firing decisions at this point are regulated by unions.
- Production and consumption take place.

Let us start with the determination of the wage contract. Since sector B firms are competing to attract workers before the workers know their skill level, they will offer the wage contract (γ, β) to solve

$$\max_{\gamma,\beta} \{\phi u(\gamma + \beta\eta) + (1 - \phi)u(\gamma + \beta\alpha)\} \tag{14}$$

subject to the nonnegative profit constraint (6) and the no-quitting constraint for the skilled workers

$$u(\gamma + \beta\eta) \geq u(A\eta - \bar{e}). \tag{15}$$

Because of the monotonicity of $u(\cdot)$, this new no-quitting constraint (15) is equivalent to the risk-neutral no-quitting constraint (5). There are two separate cases to consider. First, the no-quitting constraint may be slack, in which case the wage contract (γ, β) can be fully characterized by the first-order condition of the problem in (14) and the zero profit condition. It is immediate to see that the optimal contract gives full insurance in equilibrium to workers, so that $\beta^* = 0$. This, in turn, is consistent with the no-quitting constraint being slack if and only if:

$$\phi\eta + (1 - \phi)\alpha > A\eta - \bar{e}. \tag{16}$$

In the case where condition (16) does not hold, full insurance is not possible. Then, firms will offer wages such that skilled workers are just indifferent between quitting and staying, so

$$u(\gamma + \beta\eta) = u(A\eta - \bar{e}), \tag{17}$$

and the wage contract is determined by condition (17) and the zero profit constraint. In this case, there will only be partial insurance in equilibrium, with $\beta^* > 0$. This characterizes the equilibrium of this economy, which we describe more formally in:

Proposition 3. *There is a unique equilibrium where in sector A wages are* $w^A(\eta) = A\eta$. *In sector B, the equilibrium wage structure is characterized*

by two cutoff levels $\eta^* \equiv \frac{\overline{e}}{A-1}$ *and* $\eta^{**} \equiv \frac{\overline{e}+(1-\phi)\alpha}{A-\phi}$ *such that: (a) for* $\eta > \eta^*$ *there is no unionization and firms pay* $w^B(\eta) = \eta$, $w^B(\alpha) = \alpha$; *(b) if* $\eta < \eta^{**}$, *there is unionization with full insurance at wages* $w^B(\eta) = w^B(\alpha) = \phi\eta + (1-\phi)\alpha$; *(c) and if* $\eta^* \geq \eta \geq \eta^{**}$, *there is unionization with partial insurance, where wages are given by contract (1) with* $\beta^* = 1 - \frac{\overline{e}-(A-1)\eta}{(1-\phi)(\eta-\alpha)}$ *and* $\gamma^* = (1-\beta^*)[\phi\eta + (1-\phi)\alpha]$. *All workers with cost of schooling less than* κ^* *acquire education and enter sector A, while the rest of the workers enter sector B and join unions, where* κ^* *is defined implicitly by*

$$u(A\eta - \kappa^* - \overline{e}) = \begin{cases} u(\phi\eta + (1-\phi)\alpha - \overline{e}) & \text{if} \eta < \eta^{**} \\ \phi u(\gamma + \beta\eta) + (1-\phi)u(\gamma + \beta\alpha) & \text{if} \eta^* \geq \eta \geq \eta^{**} \\ \phi u(\eta - \overline{e}) + (1-\phi)u(\alpha - \overline{e}) & \text{if} \eta > \eta^* \end{cases}.$$

For $\eta \leq \eta^*$, *no worker quits sector B, whilst when* $\eta > \eta^*$, *skilled workers switch to sector A ex post.*

Proof. The proof is identical to that of Proposition 2. Simply note that η^* is obtained by solving the no-quitting constraint of skilled workers in sector B (17), and η^{**} is obtained by solving the condition for skilled workers not to leave their employer even when there is full insurance, (16), as an equality.□

In this economy, unions again have a useful role. Without unions, a sector B firm subject to the wage contract (1) with some degree of insurance (i.e. $\beta < 1$) would find it profitable to fire all workers who turn out to be unskilled. By imposing firing restrictions, unions eliminate this possibility. Similarly, firms have an incentive to compete to steal skilled workers from each other, as skilled workers are paid below their marginal product. Competition for skilled workers would bid their wages up to their marginal product, and insurance would disappear. Unions also prevent this type of competition by imposing hiring restrictions, such as ports-of-entry policies.

4.2.1 *Skill biased technical change*

The impact of skill-biased technical change on wage inequality is, as in the previous models, amplified by deunionization. The threshold for education, $\kappa^*(\eta)$, is again an increasing function of η. Now consider an unexpected increase in η starting from an initial value η_0 to η_1. If the rise in η is small, so that $\eta_1 \in (\eta^{**}, \eta^*)$, this will lead to a drop in the unskilled wage because the wage contract switches from perfect to partial insurance. Unions still survive, but because the cutoff $\kappa^*(\eta)$ increases, there is now ex ante deunionization. However, if the rise in η is sufficiently large so that the new value of η is

beyond η^*, then skilled workers from sector B will switch to the education-intensive sector A and there will ex post deunionization. Moreover, as in the previous subsection, when unions disappear, unskilled wages fall to their competitive wage level α.

4.2.2 *Welfare*

It is easy to extend our discussion on welfare for the training model to this insurance model. Unions are useful in this economy because they induce wage compression, thereby providing insurance to risk-averse workers. When $\eta < \eta^{**}$, the unionized economy actually achieves the first best. For higher values of η, unions are still able to improve upon the competitive economy by permitting the firm to offer some insurance. Hence, unions are efficiency-enhancing institutions which enable firms to provide credible insurance to workers. Deunionization implies that this type of insurance disappears.

5 Concluding remarks

The debate on the causes of increased inequality in the UK and the US has drawn a sharp distinction between demand-side explanations such as skill-biased technical change and increased international trade, and institutional explanations like the decline of unions and minimum wages. The timing of the changes in the wage structure and differences in cross-country patterns, such as the more limited increase in inequality in the more unionized economies, certainly suggest that the degree of unionization may be an important determinant of earnings dispersion.

In this paper, we proposed a framework that links technical change and deunionization. Our argument rests on the view that unions exist because they provide some benefits—either to the society as a whole, or simply to specific groups of workers—and they do so by imposing wage compression across workers with different skills. When the degree of skill bias of new technologies is low, the benefits provided by unions outweigh the costs of wage compression for skilled workers, who are willing to unionize. However, wage compression also makes unions vulnerable to skill-biased technical change. As the more productive employees face improved outside opportunities, wage compression becomes harder to sustain, and these workers quit unions and cause deunionization. Our approach implies that although deunionization may not be the primary cause of the surge in wage inequality, it amplifies the direct effect of skill-biased technical change by removing the wage compression imposed by unions.

Although our paper covers two of the most common views of unions in the literature (rent-seeking and efficiency-enhancing), it by no means spans

all the interesting possibilities in modelling union behavior. In particular, our model misses three important dimensions of unions. First, we used a set-up with constant returns to scale production technology, so firms make no profits in equilibrium and there are no firm rents to be extracted by the union. Second, we modeled an extreme form of collective bargaining whereby the union has all the bargaining power and imposes the wage contract to the firm. A richer framework would include the case where the union and the firm have both non-zero bargaining power. Finally, the outside option of firms was set equal to zero, which precluded a discussion of increased competition or the possibility of outsourcing or international trade changing the distribution of bargaining power. In our companion paper, (Acemoglu, Aghion, Machin and Violante, 2001) we analyze these issues theoretically and empirically, and investigate the joint effect of technical change and increased competition on deunionization.

A natural extension of our framework would allow for an endogenous technology adoption decision, which would be affected by the existence of unions in the firm. If the new technologies display larger skill bias than the old ones, unions might oppose their adoption. However, the productivity gap between unionized plants using old technologies and non-unionized plants using new technologies means that in the long run the former plants are likely to shrink, leading to a decline of unions. Moreover, firm owners and managers may install skill-biased technologies precisely in order to prevent unionization. The presumption that technology choices may affect unionization finds some empirical support from the finding that in the UK a sizeable fraction of the fall in union density is due to old unionized plants closing and new plants not recognizing a union (see Gosling and Machin, 1995). Further analysis of the two-way interaction between unionization and technology choice is a fruitful area for future research. Another important area for further study is an investigation of what type of workplace organization should replace unions. Our presumption is that different organizations are likely to complement different types of technologies, and in particular the new information technologies may call for new worker organizations rather than a total lack of organization in the new workplaces.

6 Data Appendix

Data for Figure 1 (US): Union density for the US for the years 1965, 1970, and 1975 is taken from Riddel (1993), Table 4.1, column (2), page 110. It refers to membership rates for non-agricultural male workers. The data for the missing years are obtained by linear interpolation. From 1976 to 1992 union density is taken from DiNardo, Fortin, Lemieux (1996), Table 1. The data on the 90-10 weekly wage differential for male workers are from Gottschalk (1997), Table 3, page 28.

Data for Figure 2 (UK): Union density for the UK is taken from Gosling (1998), Figure 1. It refers to union membership for male workers. The measure of wage inequality is the 90-10 percentile ratio of real male hourly earnings. The source of the data is the WIRS.

Data for Figure 3 (CANADA): Union membership for Canada refers to male non-agricultural workers and is taken from Riddel (1993), Table 4.1, column (5), page 110. The measure of wage inequality is the standard deviation of log earnings for males aged 25-58 and is taken from Baker and Solon (1999), Table 2, last column.

Data for Figure 4 (US): The data on union membership for the US are the same as in Figure 1. College enrollment data are reproduced from Gottschalk (1997), Figure 4.

Data for Figures 5,6,7 (US - Panel Study of Income Dynamics): Union membership information is available in the PSID only from the 1976 survey on. We have used all subsequent waves until 1993, the year of the last final release. Our sample comprises of all the male heads of households between 18 and 60 years old who were employed in non-agricultural sectors in each year, but not self-employed or resident in Alaska or Hawaii. We also dropped all the observations of the low-income oversample and weighted the remaining observations by the core sample family weights.

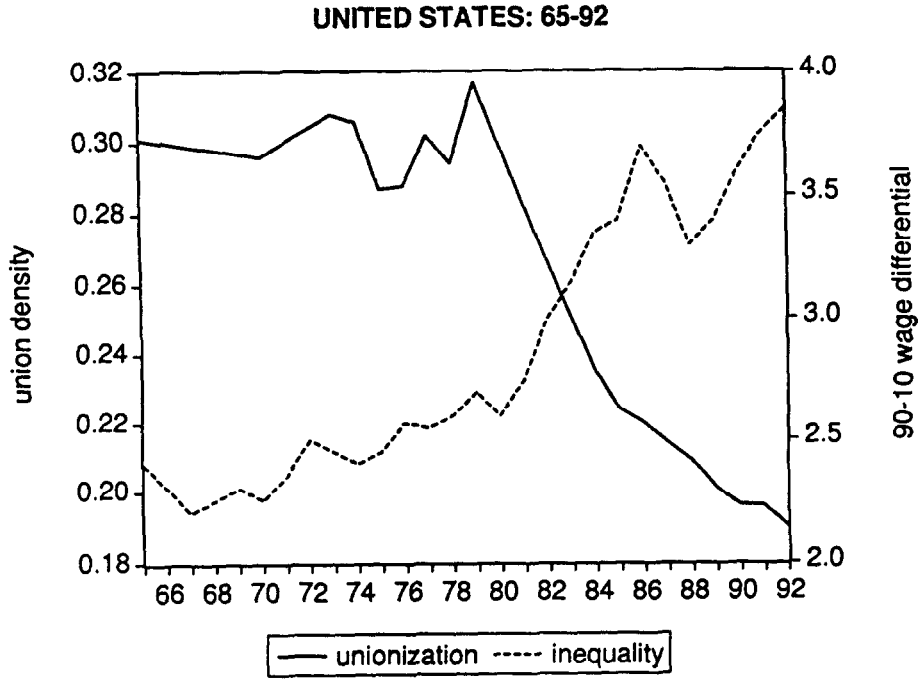

Figure 1: Unionization Rate and Wage Inequality in the US

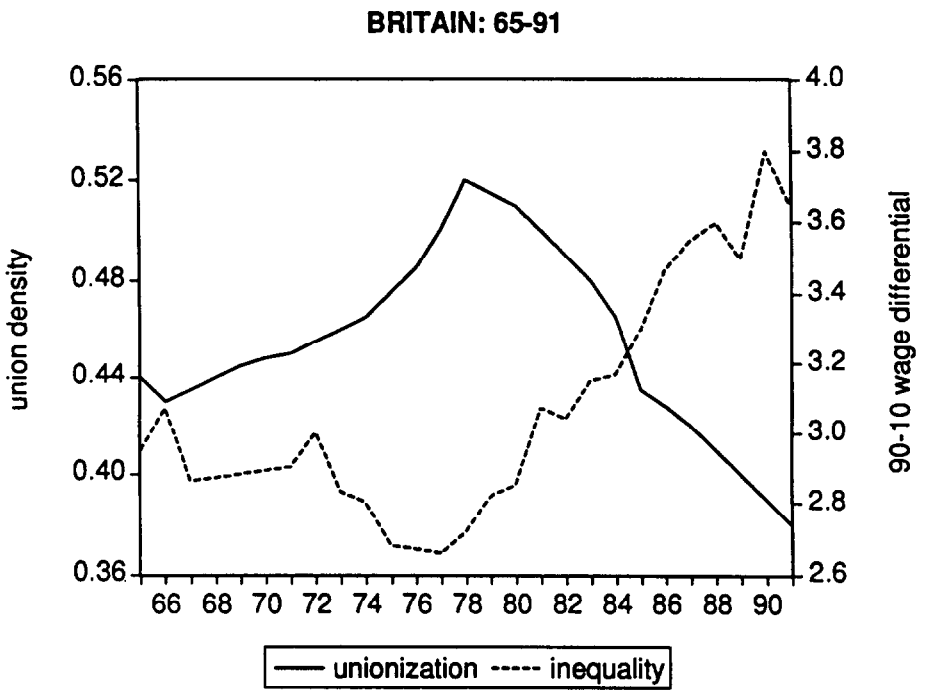

Figure 2: Unionization Rate and Wage Inequality in the UK

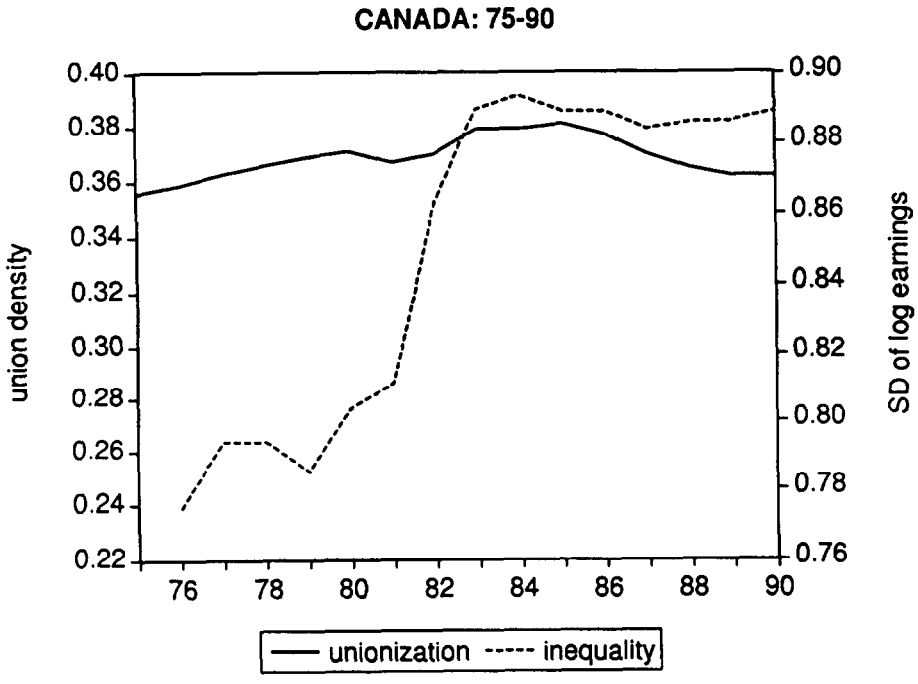

Figure 3: Unionization Rate and Wage Inequality in Canada

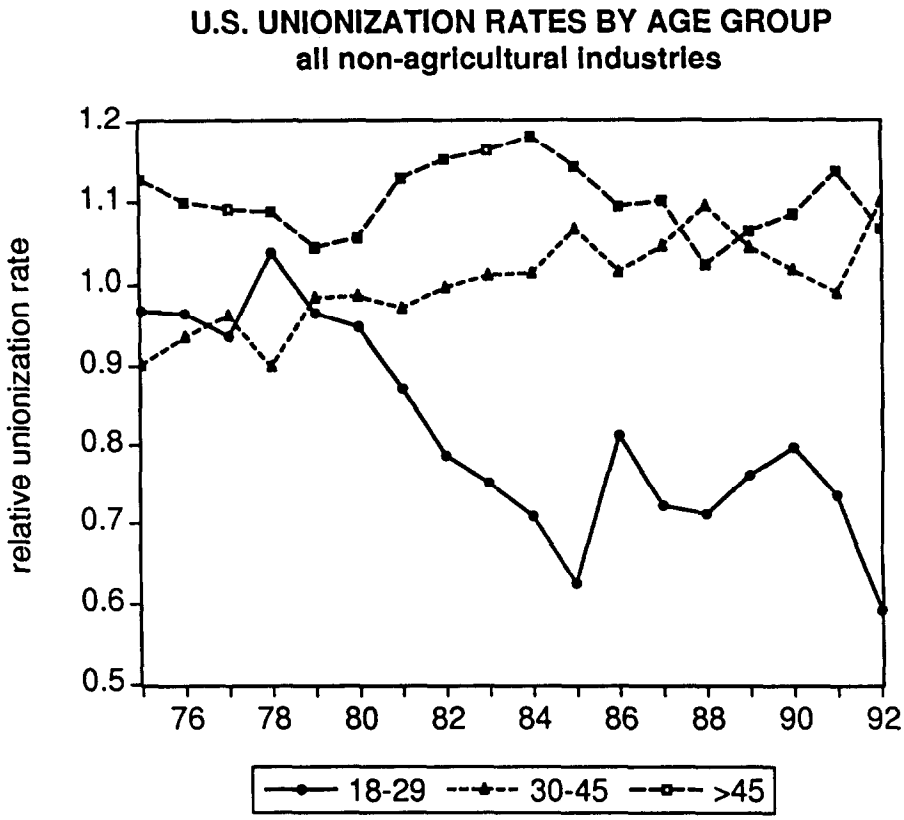

Figure 4: Unionization Rate by age group in the US - All Industries

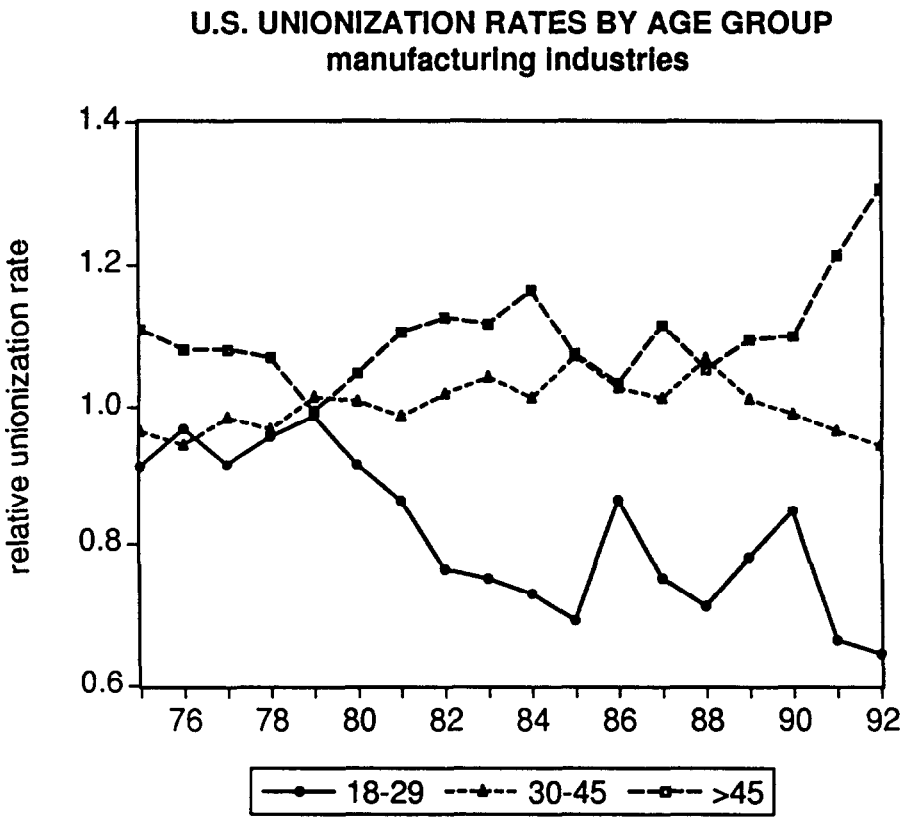

Figure 5: Unionization Rate by age group in the US - Manufacturing Industries

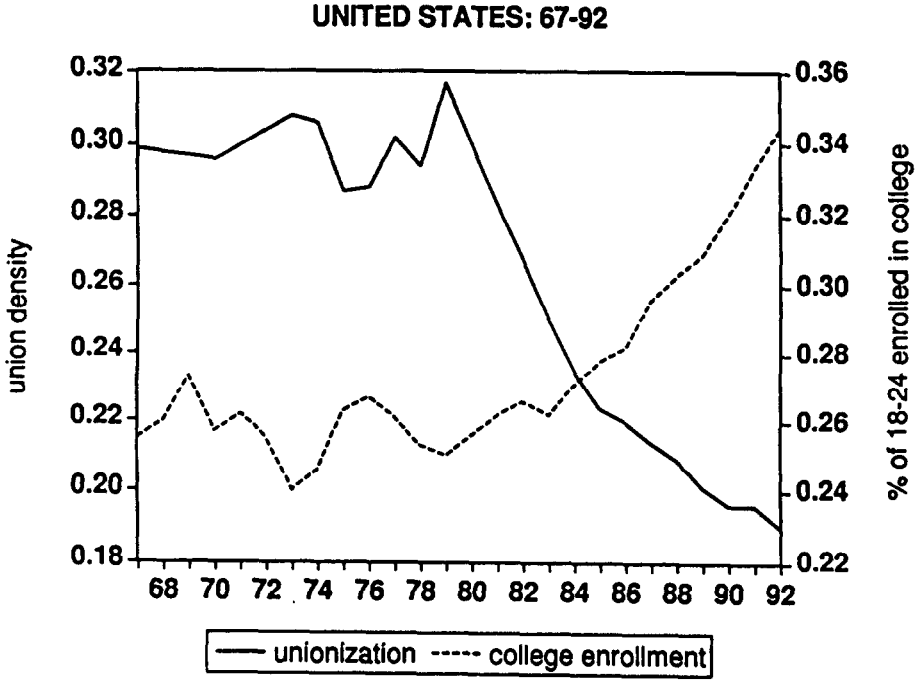

Figure 6: College Enrollment and Unionization Rate in the US

References

Acemoglu, D., (1998). Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *Quarterly Journal of Economics*, **CXIII**, 1055-1090.

Acemoglu, D., (2000). Technical Change, Inequality and the Labor Market. NBER Working Paper No. 7800.

Acemoglu, D. and Pischke, J-S., (1999a). Beyond Becker: Training in Imperfect Labor Markets. *Economic Journal Features* **109**: February 1999, F112-F142.

Acemoglu, D. and Pischke, J-S., (1999b). The Structure of Wages an Investment in General Training. *Journal of Political Economy*, **107**: June 1999, 539-572.

Acemoglu, D., Aghion, P., Machin, S., and Violante, G.L., (2001). Deunionization, mimeo UCL.

Aghion, P. and Howitt, P., (1998). *Endogenous Growth Theory,* MIT Press.

Autor, D., Katz, L., and Krueger, A., (1998). Computing Inequality: Have Computers Changed the Labor Market? *Quarterly Journal of Economics*, **CXIII**: 1169-1214.

Barron, J.M., Berger, M.C., and Black, D.A., (1997). *On-the-Job Training.* Kalamazoo: W.E. Upjohn Institute for Employment Research.

Becker, G., (1964). *Human Capital.* Chicago: The University of Chicago Press.

Berman E., Bound, J., and Machin, S., (1998). Implications of Skill-Biased Technological Change: International Evidence. *Quarterly Journal of Economics*, **113**(4): 1245-79.

Booth, A., (1991). Job-Related Formal Training: Who Receives It and What Is It Worth? *Oxford Bulletin of Economics and Statistics*, **53**: 281-294.

Booth, A., (1995). *The Economics of Trade Unions.* Cambridge, UK: Cambridge University Press.

Bernt, B., and Ragan, J.F., (1997). Have Unions Implied Growing Wage Dispersion among Young Workers?. *Journal of Labor Research*, **18**(4): 593-612.

Card, D., (1996). The Effect of Unions on the Structure of Wages: A Longitudinal Analysis. *Econometrica,* **64**: 957-979

Caselli, F., (1999). Technological Revolutions. *American Economic Review.*

DiNardo, J., Fortin, N., and Lemieux, T., (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica,* **64**: 1996, 1001- 1044.

Doeringer, P. and Piore, M., (1971). *Internal Labor Markets and Manpower Analysis.* Lexington, Mass., Heath.

Farber, H. and Krueger, A., (1992). Union Membership in the United States: the Decline Continues. NBER Working Paper #4216.

Farber, H. and Western (2000). Round Up the Usual Suspects: The Decline of Unions in the Private Sector. Princeton University, Industrial Relations Sections, Working Paper #437.

Fortin, N. and Lemieux, T., (1997). Institutional Changes and Rising Wage Inequality: Is there a Linkage?. *Journal of Economic Perspectives*, **11**:2, 75-96.

Freeman, R.B., (1988). Contraction and Expansion: The Divergence of Private Sector and Public Sector Unionism in the United States. *Journal of Economic Perspectives*, **2**:2, 63-88. Spring.

Freeman, R., (1991). How much has De-unionization Contributed to the Rise of Male Earnings Inequality? NBER Working Paper #3826.

Freeman R. and Lazear, E., (1995). An Economic Analysis of Works Councils, in Rogers,-Joel; Streeck,-Wolfgang, eds. *Works Councils: Consultation, Representation, and Cooperation in Industrial Relations.* National Bureau of Economic Research Comparative Labor Markets Series. Chicago and London: University of Chicago Press, 27-50.

Freeman, R. and Medoff, J., (1984). *What Unions Do.* Basic Books, New York.

Galor, O. and Maov, O., (1998). Ability Biased Technological Transition, Wage Inequality and Economic Growth. Brown University Mimeo.

Gosling, A., (1998). The Determination of Union Status and Wages in British Establishments. Discussion paper 491, University of Essex.

Gosling, A. and Machin, S., (1995). Trade Unions and the Dispersion of Earnings in British Establishments, 1980-90. *Oxford Bulletin of Economics and Statistics*, **57**:2, 167-84.

Gottshalk, P., (1997). Inequality, Income Growth and Mobility: The Basic Facts. *Journal of Economic Perspectives*, **11**:2, 21-40.

Green, F., (1993). The Impact of Trade Union Membership on Training in Britain. *Applied Economics*, **25**: 1033-1043.

Green, F., Machin, S., and Wilkinson, D., (1999). Trade Unions and Training Practices in British Workplaces. *Industrial and Labor Relations Review*, **52**:2, 179-195.

Hirschman, A., (1970). *Exit, Voice, and Loyalty; Responses to Decline in Firms, Organizations, and States.* Cambridge, Mass., Harvard University Press, 1970.

Hogan, C., (2001). Enforcement of Implicit Employment Contracts through Unionization. *Journal of Labor Economics*, **19**: 171- 195.

Juhn, C., Murphy, K.M., and Pierce, B., (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy*, **101**:3, 410-442.

Katz, L. and Autor, D., (1999). Changes in the Wage Structure and Earnings Inequality, Chapter 26, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, **3**: Elsevier Science.

Katz, L. and Murphy, K., (1992). Changes in Relative Wages: Supply and Demand Factors. *Quarterly Journal of Economics*, **CVII**: 35-78.

Kochan, T., Katz, H., and McKersie, R., (1986). *The Transformation of American Industrial Relations.* New York: Basic Books.

Krusell P., Ohanian, L., Ríos-Rull, V., and Violante, G., (2000). Capital-Skill

Complementarity and Inequality: A Macroeconomic Analysis, forthcoming in *Econometrica.*

Lillard, L.A. and Tan, H.W., (1992). Private Sector Training: Who Gets It and What Are Its Effects?. *Research in Labor Economics*, **13**: 1-62.

Lindbeck, A. and Snower, D., (1988). *The Insider-Outsider Theory of Employment and Unemployment.* Cambridge, Mass. and London: MIT Press, **xii**: 285.

Lipset, S.M., (1986). North American Labor Movements: A Comparative Perspective, in Lipset, Seymour Martin (editor). *Unions in Transition: Entering The Second Century,* San Francisco ICS press.

Lynch, L., (1992). Private Sector Training and the Earnings of Young Workers. *American Economic Review*, **82**: 299-312.

Malcomson, J.M., (1983). Trade Unions and Economic Efficiency. *Economic Journal*, (Conference papers Supplement), **93**: 50-65.

Neumann, G. and Rissman, E., (1984). Where have All the Union Mmbers Gone?. *Journal of Labor Economics*, **2**:2, 175-192.

Osterman, P., (1994). *Labor Economics and Industrial Relations; Markets and Institutions*, edited by Clark Kerr and Paul Staudohar, Harvard University Press, Cambridge.

Oswald, A., (1985). The Economic Theory of Trade Unions: An Introductory Survey. *Scandinavian Journal of Economics*, **87**: 160-193.

Pencavel, J., (2000). The Surprising Retreat of Union in Britain, mimeo Stanford University.

Piore, M. and Sabel, C., (1984). *The Second Industrial Divide*, Basic Books, New York.

Rees, (1962). *The Economics of Trade Unions*, Cambridge UK: Cambridge University Press.

Reynolds, L.G. (1967). *The Structure of Labor Markets: Wages and Labor Mobility in Theory and Practice*, New York, Harper.

Riddell, C., (1993). Unionization in Canada and the United States: a Tale of Two Countries in David Card and Richard Freeman (editors). *Small Differences That Matter*, National Bureau of Economic Research, The University Of Chicago Press, Chicago.

van Leeuwen, M., (1997). Trade Unions and the Provision of Welfare in the Netherlands 1910-1960. *Economic History Review*, **50**(4): 764-791.

Violante, G.L., (1999). Technological Acceleration, Skill Transferability and the Rise in Residual Inequality, mimeo, University College London.

Wallerstein, M. and Western, B., (2000). Unions in Decline? What Has Changed and Why. *Annual Review of Political Science*, **3**: 355-377.