Efficiency wages and unemployment

Moira Wilson*

1. Introduction

Efficiency wage theory—the idea that it may be efficient for firms to pay wages in excess of their market clearing levels—rests on two central propositions: (i) that there is a positive causal relationship between the wage rate paid and the productivity of labour; and (ii) that the effect of wages on productivity can be such that a drop in the wage rate paid will cause profits to fall. These propositions have been used as the basis of a series of dual labour market models in which payment of efficiency wages in one sector of the labour market may generate involuntary unemployment in equilibrium.

The purpose of this paper is to assess the relevance and explanatory power of these models from two distinct vantage points: (i) the role they ascribe to unemployment benefits; and (ii) their explanations for the observed distribution of unemployment across groups of workers. I argue that the portrayal of the relationship between unemployment benefits and the level of unemployment is problematic because it neglects the mediating effects of real world conditions of entitlement and other dynamics of unemployment. Explanations for the incidence of unemployment are also problematic: when the importance of efficiency wages as a means of recruiting, retraining or motivating workers is qualified, the link models make between the nature of jobs, the skills of workers and the incidence of efficiency wages breaks down; when the assumption of strict labour market duality is relaxed, the link made between the distribution of efficiency wages and the distribution of unemployment cannot be sustained; and when the importance of non-competing groups and discrimination is taken into account, the link made between the wage–productivity relationship of individual workers and the probability that they will experience unemployment does not hold. Dual labour market efficiency wage models, in their versions, cannot explain the incidence of unemployment as we know it.

The paper begins with a review of alternative efficiency wage explanations for wage rigidity, and the way in which these ideas, set within a dual labour market model, have been extended to a theory of unemployment.

2. Efficiency wages and wage rigidity

The efficiency wage literature stems from an analysis of the link between wages and productivity via the effect of income on nutrition and health in developing countries (Leibenstein, 1957). Economists studying labour markets in advanced capitalist countries have tended to discount the relevance of this original form,[1] but have developed the concept of a wage productivity link and the idea that it may be beneficial to employers to pay more than the absolute minimum required to attract labour. The result is a series of models that deal with different aspects of the employment relationship where wages can influence productivity: (i) recruitment of workers; (ii) worker retention; and (iii) worker motivation.[2]

2.1. Recruitment

Adverse selection. The first model of efficiency wages in recruitment presupposes that workers have reservation wages that are a function of their ability; a more able worker requires a higher wage offer to be induced to accept a given job than a less able worker. Firms have imperfect information about the ability of job applicants and so use the wage that an applicant is prepared to work for as a proxy for quality. They may be unwilling to offer lower wages because they believe that the quality of applicants will be adversely affected. Weiss (1980) holds that this can explain wage rigidity in the face of queues of unemployed lining up for available jobs.

Speed of recruitment. Efficiency wages may also occur when unfilled vacancies reduce productivity because capital equipment goes unused and production is foregone (Lang, 1991). In the speed of recruitment model, workers searching for jobs can hold more than one job offer at a time, and there is a delay between receiving a job offer and commencing employment during which further offers can be received and the original offer can be rejected. Rejection increases the duration and costs of foregone production to the firm. Assuming that financial considerations dominate the worker's decision, a higher wage offer will increase the probability that the firm's offer is that which is accepted. The prediction is that in equilibrium, a distribution of wages will exist for similar workers: firms that have a high degree of capital intensity, for whom unfilled vacancies are costly, will follow a high-efficiency wage strategy. Above-market clearing efficiency wages and involuntary unemployment will result.

2.2. Retention

Adverse selection. Firms may perceive that paying lower wages would cause their best workers, who can most easily find alternative work, to leave. Adverse selection in the 'quit response' to a lower wage may leave a firm with only its least productive workers and lower overall productivity. This may provide an explanation for the fact that firms often prefer to layoff workers rather than cut wages when they are forced to reduce their wage bill (Weiss, 1980).

[1] Stiglitz (1986); Akerloff and Yellen (1986). The relevance of the original formulation to advanced welfare economies should perhaps not be immediately dismissed. The adequacy of nutrition of children in low-income households, and the effects of child poverty on educational attainment and aspirations are still sources of concern in many advanced countries.

[2] This is the categorisation used by Layard, Nickell and Jackman (1991).

Turnover costs. When workers quit, firms incur costs of recruiting and training new workers. When firms are concerned about minimising or avoiding replacement costs associated with turnover, they may attempt to stop workers leaving by offering wages above the market clearing rate. This raises the costs of quitting to the worker in terms of both the drop in income they would incur in moving to alternative work and the increased probability that they may, in the process, suffer the unemployment produced by the existence of wages above the market clearing rate. In this model, the efficiency wage and the associated rate of unemployment will be higher the greater the costs of training workers and the higher the initial turnover propensity (Salop, 1979; Stiglitz, 1986; Layard *et al.*, 1991).

2.3. Motivation

In most jobs, workers have some degree of discretion over the effort and care they apply to their work. They can often choose, within bounds, the pace and intensity of work, and whether or not to take initiative and care or 'go the extra mile' for their employer. They can choose whether or not to take moments of leisure in work time, and, in the extreme, whether or not to hinder or sabotage production. In most firms, it is costly for employers to monitor workers constantly, and it is infeasible or costly to design a contract that pays according to the output of each individual worker (Williamson *et al.*, 1975). When entering into an employment contract, employers usually do not purchase actual production. They instead purchase workers productive potential or time on the job. Models of worker motivation look at the role that efficiency wages may play in eliciting effort from workers.

Sanctioning shirking. The shirking model of efficiency wages has as its starting points a work force that gains utility from leisure on the job and disutility from work, and a set of firms that cannot monitor workers' productivity perfectly. Shapiro and Stiglitz (1984) assume that, initially, if a worker is caught shirking they are immediately fired, but that there is no stigma attached to being fired, and the worker can easily find alternative work. In order to make being fired costly to workers and thereby discourage shirking, the firm raises the wage above that being offered by other firms. This is efficient as long as the productivity gains exceed the wage costs. If all firms in a labour market follow this strategy, wages spiral above the market clearing wage and unemployment results. Equilibrium is reached when unemployment itself is sufficiently high, and the probable duration of unemployment is sufficiently long, to deter shirking. At this point there is no need for employers to continue outbidding one another. Equilibrium unemployment plays the role of a 'worker discipline device.'

Labour extraction. Bowles' (1985) Marxist model of labour extraction is similar in structure to the shirking model. Employers seek to extract work from workers where the amount of work performed is the subject of a conflict of interests. Employers issue explicit and implicit commands and employees can resist or comply, with the threat of being fired if they are observed working below expectation. Working below expectation, however, need not entail shirking, but may lie anywhere along the effort continuum. Efficiency wages and the resulting unemployment enhances

employer power to extract labour. This is augmented by a model of wage discrimination. Worker unity increases the costs of surveillance by reducing the willingness of workers to supply employers with information about one another, and increases the possibility that non-co-operation will be in the form of a 'go-slow' or a strike. Unity declines with increased wage inequality for workers identical in their productive capacity. The need to divide and rule makes it efficient for employers to pay a wage above the market clearing rate to subgroups of workers. However, efficiency wages, unemployment and wage discrimination are not necessary outcomes: if the institutional context in which work takes place is considered fair and legitimate by workers, or they share in the gains from their labour, the underlying conflict of interests is reduced.

Eliciting care. Another variant of the efficiency wage literature proposes that different production processes are more susceptible to damage from the deliberate or accidental actions of workers than others. In this model, workers with the potential to impose high costs on their employers if they shirk or make mistakes are paid high wages, workers with little damage potential are paid low wages (Ramaswamy and Rowthorn, 1991).

In the models described above, the efficiency wage that is compatible with equilibrium unemployment is that where the decrease in effort that accompanies a wage decrease exactly offsets the associated wage savings. The elasticity of effort with respect to the wage is exactly unity—this is called the 'Solow condition' (Solow, 1979). Akerlof and Yellen (1986) point out that shirking can have damaging effects on output beyond those associated with reduced labour input. For example, reduced care or concentration can mean that machinery or equipment depreciates faster or is damaged and taken out of production for repair more frequently. Building on this idea, Ramaswamy and Rowthorn (1991) show that the effort–wage relationship in equilibrium will depend on the susceptibility to damage of the production process.   —

Sociological models of motivation. A group of sociological models of efficiency wages depart from those described above in that they do not presume that workers are concerned only with their own utility and their own wages. They consider the possibility that the utility and wage preferences of individual workers are inter-dependent, and introduce the role of social norms regarding what is considered a 'fair day's work' and a 'fair day's pay'.

(i) *Gift exchange.* Akerlof (1982) develops a model of 'gift exchange'. He describes a case study of a group of clerical workers in which it was observed that several of the workers exceed the standard rate of output set by the firm. In standard neoclassical theory, these workers should drop their production to the required rate because they receive no additional wage and have no chance of being promoted as reward. Likewise, the firm should set higher production rates for those who regularly exceed the standard, should fire or threaten to fire those who don't meet the standard, or should introduce piece rates and pay by output. He maintains that the explanation for the failure of both parties to act 'rationally' is sociological: workers develop 'sentiment' for the firm and for one another over time. Output in

excess of the required minimum produced by the workers is a 'donation of good will and effort' in return for which they are given a 'fair wage', and tolerance of variability in ability to meet the set standard. If the firm increased the standard rate, or shed those workers who could not meet the standard, this would be considered unfair by the group. Morale and 'sentiment' for the firm would suffer, the gifting would cease, and productivity would be reduced. For this reason, minimum work standards are rarely set at the highest possible limit and wages are rarely set at the lowest acceptable level. Wages in excess of their market clearing levels are efficient for the firm. Notions of what constitutes a fair exchange of gifts are based on historical effort and wage norms, customary comparisons with reference groups, and the level of unemployment benefits and the incidence of unemployment among reference group members. Their evolution and enforcement is conditional on the existence of stable work groups.

(ii) *Fairness and productivity enhancing games.* Akerlof (1984) explores the sociology of the behaviour of another work group, this time male production workers who make a game of trying to better each other's daily output. Akerlof argues that the workers' perception of the wages paid as fair underlies their willingness to engage in a game that makes their work fun by relieving boredom *and* is of benefit to their employer: 'If wages are sufficiently low, workers will feel unfairly treated. Such unfair treatment will take the fun out of playing a game whose results benefit the firm. . . . According to this view, in the case where workers have animosity towards their employer, higher wages will cause workers to feel less badly about relieving their boredom by playing a game which yields a surplus to the firm. Or, alternatively, if workers have loyalty to their employer, low wages will cause workers to feel less badly about playing a game which fails to benefit the firm' (Akerlof 1984, pp. 81–82).

(iii) *Fair wage relativities and reference wages.* In subsequent papers, Yellen and Akerlof formalise the role of reference groups and the sense of relative disadvantage (Akerlof and Yellen, 1988; Yellen and Akerlof, 1990). According to their model, where the effort of groups of workers is in part dependent on their view of the fairness of their wage in relation to that of other workers, it is to the advantage of firms that are compelled by market forces to pay high wages to some employees to pay higher than market clearing wages to lower paid employees. Yellen and Akerlof suggest that this can explain why industries that have high wages for one occupational group tend to have relatively high wages for other occupational groups, an observation that the shirking models and replacement cost models have difficulty explaining.

While Akerlof (1982) allowed the perceptions of a 'fair wage' to depend on past wages, this is absent from the fair-wage relativity model of Yellen and Akerlof (1990). Wadhwani and Wall (1991) re-emphasise the role of past wages and past wage relativities and the processes of adaptation: 'people become accustomed to a certain state, good or bad, and therefore tend to be influenced strongly by events that are better or worse than normal' (Wadhwani and Wall, 1991, p. 544). There is some evidence that the role of past wages as a reference for judging the fairness of wage adjustments distinguishes between norms of fairness for incumbent workers and for new workers. Survey responses show that where workers are available at

lower wages than an incumbent worker is paid, a wage reduction for the incumbent worker is considered unfair; a lower wage for a replacement worker when the incumbent leaves, however, is considered acceptable (Kahneman *et al.*, 1986).[1] Okun suggests that the wage stickiness implied by these norms applies at all levels of the labour market (Okun, 1980).

Perceptions of fairness are not instantaneously sensitive to aggregate demand shocks. This, and the sanctions that workers can apply if they perceive a wage reduction to be unfair, may provide a partial explanation for cyclical unemployment. Kahneman *et al.* (1986) note that while in some cases concern about fairness can permanently prevent the market from clearing,

[i] [I]n other situations, the reluctance of firms to impose terms that can be perceived as unfair acts as a friction-like factor. The process of reaching equilibrium can be slowed down if no firm wants to be seen as a leader in moving to exploit changing market conditions. In some instances an initially unfair practice . . . may spread slowly until it evolves into a new norm—and is no longer unfair. . . . In all these cases, perceptions of transactors' entitlements affect the substantive outcomes after changes, altering or preventing the equilibrium predicted by an analysis that omits fairness as a factor. In addition, considerations of fairness can affect the form rather than the substance of price or wage setting. Judgements of fairness are susceptible to substantial framing effects. . . firms have an incentive to frame the terms of exchange so as to make them appear 'fair'. (Kahneman *et al.*, 1986, p. 740)

Rent sharing. Reference group comparison can also generate efficiency wages where workers perceive there to be an unfair disparity between their earnings and those of the firm they work for. Rent sharing efficiency wages may be paid where firms perceive that workers will shirk if they do not share in some part of the profits of the firm, or where rent sharing can improve morale, motivation and loyalty. This may explain why firms with high profit margins tend to pay high wages (Akerlof and Yellen, 1988). An alternative explanation for rent sharing is that where profit margins are high, managers can spend some of the derived rent on high wages so as to minimise industrial tensions and make their own jobs easier. This explanation involves 'rent capture' by managers, rather than an active efficiency wage strategy on the part of the firm shareholders (Layard *et al.*, 1991).

3. Dual labour markets, efficiency wages and unemployment

Efficiency wage theory has been extended to form the basis of a series of models of involuntary unemployment. In these models, the wage–productivity relationship is important to varying degrees: for some employers it is worthwhile to use efficiency wage strategies to promote productivity, for others it is not. Theorists adopt the concept of a 'dual labour market' and the distinction between 'primary' and 'secondary' labour markets coined by Doeringer and Piore (1971) as a framework for this diversity. Set within a dual labour market, it is argued, efficiency wage theory can provide insights into the distribution and dynamics of unemployment.

The statement of dual labour market theory that characterises the use of the concept in the efficiency wage literature is this:

[1] It is not clear from their study whether the same result would apply if the lower waged newcomer were to work alongside incumbent workers paid at the old higher rate. Nor is it clear whether it is considered fair to fire the incumbent high-paid worker in order to take on the new low-paid worker.

Jobs in the primary market possess several of the following characteristics: high wages, good working conditions, employment stability, chances of advancement, equity and due process in the administration of work rules. Jobs in the secondary market, in contrast, tend to have low wages, few fringe benefits, poor working conditions, high labour turnover, little chance of advancement and often arbitrary and capricious supervision. (Doeringer and Piore, 1971, p. 165)

Movement between the two sectors is limited. Entry into the primary labour market is restricted by institutional rules for rationing 'good' jobs and the feedback effects of secondary employment on the actual or ascribed characteristics of workers who spend time in it.

Efficiency wage models of unemployment generally delineate labour market duality as follows: Where the wage–productivity link is important and firms use wages to elicit effort or to recruit or retain workers, the wage is above the market clearing wage. Because of this, jobs are rationed—not all those who seek such jobs can find them. This is the primary labour market. Where the intensity of work can easily be controlled and monitored, or the wage productivity link is not strong, there is no need for firms to offer efficiency wages, and the wage paid equals the market clearing wage: 'anyone can obtain a job in this sector, although it might be at lower pay' (Akerlof and Yellen, 1986, p. 3). This is the secondary labour market.

If workers who seek employment in the primary labour market are unable or unwilling to search while employed in the secondary labour market, a queue for primary labour market jobs forms. The unemployment associated with this queue is involuntary in the sense that those in the queue are willing to work at the going wage in the primary labour market, but cannot find work there. Relative to employment in the secondary labour market, unemployment may improve a worker's chances of being offered primary sector employment if it enables more effective job search. Alternatively, if primary sector employers perceive willingness to accept a secondary sector job as a signal of low worker quality and low employment aspirations, or willingness to job hop out of the secondary sector as a signal of high turnover propensity, then remaining unemployed is a rational strategy. Empirically, few people who lose primary sector jobs accept 'stopgap' jobs in the secondary sector. Bulow and Summers (1986) suggest that this is evidence that signalling considerations are important.

What follows is a critical review of the role ascribed to unemployment benefits in these models, and their predictions concerning the distribution of unemployment experiences across different labour market groups.

4. Unemployment benefits

Central to efficiency wage models is the idea that a positive rate of unemployment in equilibrium is functional to primary sector employers. It allows them to elicit effort, to recruit rapidly and from a large pool of applicants, or to retain workers without the need to raise wages above those offered by other employers; the probable duration of unemployment in itself is sufficient to discourage workers from shirking or quitting, and the level of unemployment in itself is sufficient to generate a large number of applicants.

In this context, an increase in the level of unemployment benefits changes the efficiency gains to employers associated with a given level of unemployment: (i) In the shirking model and turnover dual labour market models, it will reduce the costs of being fired or leaving a job associated with any given level of unemployment, raise the efficiency wages required to elicit effort or discourage quits, and in turn raise the equilibrium level of unemployment (Bulow and Summers, 1986; Stiglitz, 1986; Layard et al., 1991). (ii) In the speed-of-recruitment model, it may reduce the search intensity of the unemployed, raise the efficiency wage required to attract a given number of applicants and secure rapid acceptance, and again raise the equilibrium level of unemployment. (iii) In the adverse selection recruitment model, it may reduce the search intensity of lower productivity workers and thereby reduce the efficiency wage required to attract a pool of applicants of a given quality and thereby reduce equilibrium unemployment (Stiglitz, 1986). (iv) In the gift-exchange and fair-wage models, it will reduce the utility of any given exchange of gifts to a worker, and, where their perception of a fair wage is determined in part by reference to the incomes of unemployed workers, raise the wage they require in order to supply a given amount of effort.

There are two major problems with these portrayals. The first is the level of abstraction from the rules and conditions of benefit receipt that apply in most real world systems. The second is the unclear relationship between the demand side efficiency wage effects of benefit levels on unemployment and the supply side financial incentive effects of benefit levels on unemployment.

4.1. Real world conditions of entitlement

Atkinson and Micklewright (1991) point out that the predictions of efficiency wage models are based on very simplistic and often highly unrealistic assumptions about the way unemployment benefit and unemployment insurance systems are administered. Not all unemployed workers receive benefits. In particular, a worker who is fired for misconduct or for shirking, or a worker who leaves their job voluntarily is often disqualified from receipt of benefits. A change in the level of unemployment benefits will have no impact on the level of efficiency wages required to elicit effort or retain workers where this is the case.[1]

Atkinson (1992) demonstrates that modifying the shirking model of unemployment to account for the administrative realities of unemployment insurance regimes produces vastly different predictions from the standard model. Assuming that employment in the secondary sector labour market is not covered by unemployment insurance contributions, an increase in the level of unemployment insurance payments increases the relative value of primary sector employment by reducing the costs of termination relative to those in the secondary sector. In this way unemployment insurance operates 'as a subsidy on the uncertain employment in the primary sector' (Atkinson, 1992, p. 102) and the increase in the value of unemployment insurance to primary sector workers lowers primary sector wage required to sanction shirking. Employment in the primary sector is increased.

---

[1] Atkinson (1992) observes that where monitoring or enforcement of these conditions is less than perfect, the outcome will be somewhat different. Where knowledge of these rules is less than perfect, there may still be some impact.

Within the same framework, an increase in the stringent of job search require-
ments and mandatory job acceptance rules will lower the value of unemployment
insurance as insurance against job termination to primary sector workers. This will
have the seemingly paradoxical effect of increasing the wage required for no
shirking, and raising unemployment. If conditions of entitlement mean that
unemployed workers in receipt of unemployment benefit or unemployment insur-
ance are not at liberty to decide the intensity or the direction of their job search, this
may also mediate against the positive effect of a benefit increase on primary sector
employment possible in the adverse selection recruitment model.

4.2. Financial incentives

The dual labour market model embodies a second more direct route through which
the availability and level of unemployment benefits can affect unemployment.
Unemployment benefits can provide incentives for unemployed primary sector
workers to wait in the queue for primary sector jobs rather than take up jobs in the
secondary sector, and can induce unemployed secondary sector workers to join the
queue for primary sector work rather than seek re-employment in the secondary
labour market. Setting aside for a moment the fact these incentive effects would also
be mediated by conditions that determine eligibility for benefits,[1] this second route
introduces an essential ambiguity which has not been addressed in the literature. If
higher benefit levels simultaneously reduce the effectiveness of the existing rate of
unemployment in helping firms recruit, retain or motivate workers, and increase the
rate of unemployment to a new level, the net effect of the two processes is not clear;
it is possible that the increase in the level of unemployment via the incentive effect
is sufficient to maintain levels of motivation without the necessity for an increase in
primary sector efficiency wages and a reduction in primary sector employment.

This underlines one of the problems with the efficiency wage dual labour market
models: how do they relate to other forms and other dynamics of unemployment?
In particular, how do they relate to the supply side of the labour market? If
unemployment increases because of an exogenous supply shock, the logic of, for
example, the shirking model should predict a reduction in the wages required to
motivate primary sector workers, and a reduction in unemployment. But this is the
very process that we do not see, and the absence of which efficiency wage theories
set out to explain.

The response to this problem may be that for unemployment to provide a credible
discipline to those in employment, or for the unemployed to constitute a relevant
reference group for those who are in employment, those who are unemployed must
be similar to those who are in work; workers must not perceive that if they were to
become unemployed, they would be able to find alternative work easily. If the
unemployment generated by financial incentives associated with benefits (or by
supply shocks) is concentrated on groups of workers who are unlike most primary
sector workers, then an increase in efficiency wages and contraction of primary

---

[1] Atkinson demonstrates that the commonly held view of the role that unemployment insurance has on
reservation wages is also mediated by job search and acceptance rules: 'If there is a positive probability
that benefit will be terminated, either because entitlement is exhausted or on the grounds that secondary
employment is considered a "suitable" alternative, then there is no disincentive even with 100%
replacement of net earnings (Atkinson, 1992, p. 100).

sector employment may still be required to generate a group of unemployed that can maintain primary sector incentives not to shirk or to quit, or to equilibrate their sense of what is a fair wage. And yet if this is the case, how do we reconcile the fact that those who suffer unemployment tend to be quite unlike those in primary sector employment? They are more likely to have low skill and education levels, more likely to be young, more likely to be old, and more likely to be members of disadvantaged ethnic minorities.

Dickens and Lang (1988) suggest that one of the problems with efficiency wage models of unemployment is their failure to take account of frictional unemployment, associated with 'flux and uncertainty in the secondary sector', which was the form of unemployment originally emphasised by Doeringer and Piore (1971). They advocate development of an efficiency wage model where this form of unemployment co-exists with 'wait' unemployment for primary sector work:

> If there is sufficient advantage to being unemployed while searching, some workers will prefer to remain unemployed while seeking high-wage employment. So there will be some wait unemployment. Layoff from the primary sector will be likely to result in wait unemployment since such workers will have more unemployment insurance and accumulated assets which will make waiting more affordable. . . .Workers in the secondary sector would suffer considerable frictional unemployment as the market adjusted to continual shifts in the level and distribution of demand. One of the remaining challenges for segmented labour market theorists is to develop such a model more fully and to subject it to empirical scrutiny. (Dickens and Lang, 1988, p. 132)

It is still, however, difficult to reconcile predicted and actual unemployment incidence in the preliminary framework that they set out. This is illustrated if we consider long-term unemployment. By implication from the passage above, wait unemployment spells will tend to be longer than frictional unemployment spells, and waiting will be more worthwhile and affordable for primary sector workers. And yet the long-term unemployed tend to have even less in common with primary sector workers than the short-term unemployed.

5. The incidence of unemployment

Reviewers of efficiency wage models claim that they are invaluable as explanations not only for the existence and persistence of unemployment but also for its observed distribution across groups (Stiglitz, 1986; Akerlof and Yellen, 1986). Examining the predictions of the models is a useful vantage point from which to assess their relevance.

Efficiency wage models of unemployment in a dual labour market say that where efficiency wages are paid, queues for jobs serve as an equilibriating force and involuntary unemployment exists in labour market equilibrium. On a very simplistic level, therefore, the distribution of efficiency wages can perhaps tell us something about the distribution of this equilibrating unemployment: involuntary unemployment will comprise groups that are difficult or costly to monitor, groups that can cause damage to equipment or to output if they are not encouraged to work with care, groups that work in capital intensive or high value of product industries where

the costs of unfilled vacancies are high, groups that work in profitable industries where they can extract rents on the grounds of fairness or on the threat of industrial disharmony, groups that are highly sought after employees, and are the hardest to retain, groups that are highly skilled and therefore the most costly to train and replace, groups that have a gift exchange relationship with their employers and can impose costs on their employers if they contravene standards of fairness, and groups that work alongside others that are highly paid and can impose costs on their employers if they contravene standards of fairness.

This approach is common in the literature. The speed of recruitment model is held to provide a partial explanation for lower rates of unemployment among white-collar workers who tend to work in jobs that have relatively low capital intensity and therefore low costs associated with unfilled vacancies (Lang, 1991, p. 189). Akerlof and Yellen (1986) observe that the shirking model predicts that highly skilled professional workers, who have a high degree of self-direction and autonomy in their work and are therefore extremely hard to monitor, should have efficiency wages and high rates of involuntary unemployment. They argue that the low rates of actual unemployment for these groups implies that they get less utility than other workers from shirking. Yellen and Akerlof (1990) claim that only their efficiency wage model can predict efficiency wages for the low paid and thereby explain the association of unemployment with low skill levels and low pay.[1]

This view of the distribution of efficiency wages and the associated involuntary unemployment can be highly misleading. There are three reasons: (i) the wage level is not the only instrument available to firms to recruit, retrain and motivate workers and once the importance of efficiency wages has been qualified, it becomes apparent that *there is no necessary link between the nature of jobs, the skills of workers and the incidence of efficiency wages*; (ii) dual labour markets are an analytical tool not a reality, and once the assumption of duality is relaxed and the labour market is viewed as segmented along a continuum, the nature and the function of the queue of those who are involuntarily unemployed looks quite different—*there is no necessary link between the distribution of efficiency wages and the distribution of unemployment*; and (iii) efficiency wage models tend to neglect the role of discrimination and the idea that there are non-competing groups separated by the internalisation of labour markets—*being in the queue of unemployed workers, and ranking in the queue, has little to do with the wage–productivity relationship of individual workers*. I take each of these points in turn.

5.1. Qualifying the importance of efficiency wages
Efficiency wages are not the only strategy open to a firm that seeks to attach workers to it and motivate them. Indeed, the theory of internal labour markets that

---

[1] This particular model has problems in addition to those outlined below. Yellen and Akerlof assume that high and low-wage workers are reference groups for one another, and that the perceived fair wage of each is some weighted average of their own market clearing wage and the wage paid to the other group. By definition, the fair wage of the high-paid group cannot be greater than their market clearing wage, they are fully employed, and there is no mechanism by which closing the gap between high- and low-paid workers results in the exertion of fair-wage pressure by them. Empirically, however, fair wages norms tend to be manifest in historical wage relativities to workers both up and down the wage distribution, and wage compression from below may result in wage pressure from the higher paid who seek to maintain these relativities.

underpinned the dual labour market hypothesis of Doeringer and Piore (1971), and has been built on by subsequent theorists, suggests that a range of recruitment, retention and motivation imperatives and strategies exist. The importance and relevance of efficiency wages must be qualified.

Other costs of being fired. In their original formulation of the shirking model, Shapiro and Stiglitz (1984) note that dismissal can be costly to a worker even in the absence of income loss associated with an efficiency wage. Fired workers may lose the ability to use and be compensated for skills and knowledge that are specific to the firm in which they work, the costs of searching for and moving into an alternative job may be significant, and the stigmatising effect of dismissal may down-grade their opportunities (Doeringer and Piore, 1971). The threat of dismissal may be an effective discipline device without efficiency wages and efficiency wage related unemployment.

Negative and positive motivation strategies. Moreover, motivation strategies do not necessarily require reliance on the threat of dismissal. Firms have at their disposal many strategies for motivating their workers. Green and Weisskopf (1990) point out that the association of unemployment as a discipline device with primary sector employment used in the dual labour market model of Bulow and Summers (1986) is at odds with studies that suggest that in fact the threat of dismissal tends to be more likely to be used in the secondary labour market, and more bureaucratic forms of 'positive motivation' and control tend to be used in the primary labour market. For many workers, the threat of dismissal is not credible; workers and firms have too much invested in one another for firing to be a realistic sanction. Thus 'While the ultimate threat of dismissal is never completely absent, alternative negative sanctions are frequently applied—such as the withholding of wage increases, denial of promotion, imposition of fines, or sometimes demotion to less skilled jobs. Workers may also be motivated to work hard by means of various positive incentives, such as regular wage and benefit increases, favourable working conditions, and/or a corporate ideology that promotes high worker morale and calls forth loyalty to the company's aims' (Green and Weisskopf, 1990, p. 241).

Self enforcing implicit contracts. Similar qualifying statements apply to the turnover model. The very factors that make turnover costly for firms—hiring costs and the costs of firm specific training—also make turnover costly for workers, but in the form of search costs and loss of firm-specific skills. In the language of the literature on self-enforcing implicit contracts, unemployment is only one of several possible sources of the 'surplus' that makes it beneficial for both workers and firms to honour their sides of an implicit agreement and foster a long-term relationship (Carmichael, 1989). Other sources of surplus, such as the costs of job changing or worker replacement, firm-specific skills and knowledge, loss of reputation associated with being fired or firing workers without just cause, the existence of third parties that will ensure compensation is paid upon separation, or a sense of moral justice or loyalty, give employers the incentive and the means to retain labour.

Costless asset formation. Where workers enjoy privileges that make it costly for them to exit from a firm, these are not necessarily in the form of efficiency wages, and

need not be costly to the firm; Doeringer (1986) emphasises the role of 'costless asset formation' in binding workers and firms together. Examples of costless assets are skills that are acquired on the job but do not involve the firm actively investing in training (learning by osmosis), or the formation of friendships at the workplace. One of the foundations of the original dual labour model was the observed use of job ladders and internal promotion (tournaments in Carmichael's terms) as a means of motivating and retaining workers and, importantly, as a means of ensuring that workers felt secure enough to share their skills and knowledge with other workers and thus facilitate relatively costless on the job training for firms.

The bonding critique. Critics of the shirking and retention efficiency wage models argue that an alternative to efficiency wages is to require a worker to pay an entry fee or 'post a bond' upon commencing employment which would be forfeited if they were found shirking or quit, or repaid at the end of their working life. This would solve the problem of discouraging shirking or quitting without generating unemployment. Shapiro and Stiglitz (1984) argue that bonding is not observed in practice because unemployed workers do not have sufficient funds themselves and there is no capital market that will allow them to borrow for this purpose, and because there is a moral hazard problem associated with the incentive that employers would have to wrongly accuse an employee of shirking, fire them, and steal their bond. Carmichael (1990) argues that imperfect capital markets cannot explain the absence of bonds, and that when unemployment is not the only source of surplus from an employment relationship, and particularly when employers are concerned about their reputations, they have no incentive to steal their workers' bonds. Relatively low wages at entry and wage progression with age may be interpreted as serving the purpose of an entry fee, although Carmichael argues that rising age wage profiles still entail efficiency wages. Theoretical debate over the issue of bonding has been protracted. Theorists on each side of the debate have claimed that the absence of bonds is evidence for and against the existence of shirking preventative efficiency wages.[1]

Recruitment in an extended internal labour market. There are also alternative strategies to efficiency wages for filling vacancies speedily and attracting applicants of the desired quality. In his formulation of the model, Lang notes that the importance of efficiency wages needs to be qualified: 'In no way does it preclude the use of other mechanisms designed to minimise the occurrence and cost of vacancies. Firms may overoffer, hoard labour, choose technologies that are relatively insensitive to vacancies, and choose a scale of production designed to minimise such costs (1991, pp. 187–188). Another strategy firms may use to fill vacancies quickly is to recruit from among the friends and relatives of their workforce. Use of this 'extended internal labour market' may also serve as a control on the quality of labour attracted—firms can select only the best workers as contacts for word-of-mouth recruitment and, by using the interest of the contact in preserving their own reputation as a good contact and a good worker, reduce the effort they need to expend to ensure that the new worker performs (Manwaring, 1984). There is

[1] See Carmichael (1990, pp. 283–284) and Lang and Kahn (1990, pp. 297–299).

evidence that word-of-mouth recruitment can account for a large proportion of total hires (Montgomery, 1991; Reid, 1972).

Non-wage recruitment strategies. Moreover, because employment relationships in the internal labour markets of the primary sector tend to be long term, reputation as a good employer, and the ability of a firm to provide job security, training, child-care facilities or opportunities for advancement may be as important as the starting wage to prospective applicants; the decisions workers make are not solely driven by comparison of the wages immediately available, and high starting wages are not necessarily coupled with good working conditions or good 'prospects'. As an explanation for unemployment, the plausibility of the recruitment type models has its limits: 'It is hard to believe, for example, that the dramatic increase in unemployment in a recession is due to an increase in the optimal queue length for those firms that are still hiring workers' (Carmichael, 1990, p. 286). Again, if there is some optimal queue length, an exogenously driven increase in unemployment should have the effect of reducing the efficiency wage required to attain it, and reducing unemployment, but this is the very process that is absent.

Fairness and resistance or re-evaluation. We are left with the sociological models. Yellen and Akerlof (1990) note that norms of fairness can be adapted so that what is considered a fair exchange can be up or downgraded. What is considered 'fair' can be redefined. Yellen and Akerlof (1990) note that where workers receive less than what they perceive to be a fair wage for their labour, they may (i) reduce their level of effort, (ii) re-evaluate their perception of the amount of effort they put into their work, or (iii) re-evaluate their perception of the worth of that effort. This seems to be an important qualification to their models, which only look at (i), changes in actual effort. If some groups reduce their actual effort in response to a drop in wages below those previously considered fair, and other groups down-grade their perception of what is a fair exchange and accommodate the new wage rate without reducing their effort, there is no necessary correspondence between the existence of stable work groups that form views about what is fair, and efficiency wages.[1] Redefinition rather than resistance may be the strategy or the only option for groups who lack alternatives or who lack individual or group bargaining power. This may be rooted in labour market discrimination, or in relatively autonomous sociological forces such as norms of secondary income earner status for women (Humphries and Rubery, 1984) or the lack of an established sense of rights for migrants or disadvantaged minorities.

Group versus individual bargaining power. A final qualification applies to the efficiency wage literature in general: the models present wage rigidity as the result of employer strategies only. Furthermore, with the exception of the gift exchange and Marxian variants, the models tend to focus only on employer response to the costs that *individuals* can impose by withdrawing effort or by quitting. The existence of *group* bargaining power and collective action, such as working to rule or taking strike

---

[1] For example, historically the process of downgrading work that has been feminised relied on the ability to simultaneously reduce pay and reduce perceptions of worth.

action, can magnify these costs exponentially. For example, '[t]he penalty from quits by individual workers with specific skills is confined to the replacement costs of those workers, whereas collective quits impose further costs of replacing that part of the firm's on-the-job training capability that is embodied in the incumbent workforce' (Doeringer, 1986, p. 49). Group bargaining power can exaggerate or negate the wage–productivity relationship predicted on the basis of a given set of job characteristics, and while bargaining power may be in part determined by the same forces as the wage–productivity relationship, it is in part determined by quite independent historical, social and political forces. Worker strategies in themselves cast quite a different light on the causes and consequences of wage rigidity.

Together, these qualifications dispel any straightforward or automatic relationship between the nature of jobs, the skills of workers and the distribution of efficiency wages. Equally, I will argue, we cannot infer the distribution of efficiency wages by working backwards from the distribution of unemployment—the incidence of efficiency wages may bear no direct relationship to the incidence of efficiency wage related unemployment.

5.2. Relaxing the assumption of duality

Relaxing the assumption that the labour market is characterised by strict duality, and viewing the labour market as segmented by labour market internalisation, but along a continuum rather than into two polar extremes,[1] it can be argued that efficiency wage-related unemployment is not necessarily borne by workers who are similar to those who receive efficiency wages.

If jobs are viewed as falling into labour market segments along a continuum that represents some measure of job quality, the assumption that people who lose jobs in the primary labour market segments of the continuum choose between employment in the secondary labour market segments or queuing in unemployment to reenter employment in the primary segments must also be relaxed: not all jobs in the primary labour market are the same, when a worker is displaced from employment in one segment of the primary labour market they may seek and quickly find re-employment in a lower level job. This may be a permanent shift, it may be a way of carrying out further job search without incurring the scarring effects of either secondary sector employment or prolonged unemployment (although to the extent that employers perceive willingness to job hop as a sign of poor stability, some scarring effects may remain), or it may be a means of gaining access to higher level jobs within the same internal labour market. In turn, the assumption that primary sector employers only hire from the pool of unemployed workers must also be relaxed.

Layard, Nickell and Jackman summarise the implications of this process as follows:

One obvious question is, Do efficiency wages help us to explain the occupational structure of unemployment? The answer is this. Efficiency wages explain why there are job queues: employers find it in their interest to pay workers more than their expected wage outside. This applies at least as much to skilled as to other workers. However, skilled workers can usually get a less skilled job but at a lower wage. Thus in equilibrium we see less unemployment for

---

[1] For a discussion of 'heuristic' duality, see Ryan (1981).

skilled workers: they simply experience job queues for skilled jobs but not for unskilled. For unskilled workers, however, there may be 'nowhere to go'. When their employers offer them a premium, this implies that their alternative involves either a chance of unemployment or, *if* there is a secondary sector, a job there at a lower wage which they may not be willing to accept. (1991, p. 168)

What is not said, but follows from the same line of reasoning, is that the groups who suffer unemployment are not necessarily those who, when they are in work, receive more than market clearing wages. If skilled workers can always get jobs that require fewer skills than they possess ahead of less skilled workers, those who are unemployed at the end of the process of shunting less skilled (and perhaps also less experienced and less advantaged) workers down the job hierarchy may have very little in common with those who receive efficiency wages that cause the unemployment in the first place.[1]

There is a fundamental tension between the equilibrating role attributed to unemployment and the idea that there is a ranking and shunting process that determines incidence and duration of unemployment. If the unemployed are the least cost effective in terms of their wage–productivity relationship, or the least skilled, advantaged, or experienced, then how can they provide (depending on the model) the 'working discipline', or the discouragement of quitting, or the optimal queue quality, or the equilibrating reference group comparison required? If currently employed workers would, by definition, be ahead of those currently unemployed in the queue for jobs were they to be fired or were they to quit, then the length of the queue cannot equilibrate efficiency wages.[2]

This dilemma is the same as that posed above in regard to the incentive effects of benefits. I would suggest that it can only be resolved by accepting that there may be an asymmetry in the effects of the wage–productivity relationship: employer concern about the productivity implications of a drop in wages may explain wage rigidity in the face of mass unemployment; but the idea that employers generate unemployment by entering into an efficiency wage competition with other employers, stopping only when mass unemployment itself is sufficient to ensure the recruitment, retention or motivation outcomes they seek, contradicts what we know about the incidence of unemployment and the sorting and ranking processes that go on in the queue for jobs.

There may also be asymmetry in the underlying wage–productivity relationship. In the literature on perceptions of fairness, experiments suggest that if wages fall below levels previously considered fair, actual effort tends to fall; if wages rise, the worker's perception of the effort they put into their work and the value of their effort

---

[1] There may be limitations to this dynamic if employers perceive there to be potential productivity losses associated with hiring a worker who is overqualified for the job at hand. For example, employers may expect that the worker will be discontented, have trouble co-operating with other workers, or leave at the first opportunity when a better job opportunity arises. There may be a hidden discipline effect associated with downward occupational mobility—intense competition for 'good jobs' may occur without necessarily high rates of unemployment among applicants and the number of applicants may be the mechanism by which incumbent workers are made aware of the difficulty they would have gaining re-employment at the level of their current job. Downward occupational mobility may also provide a hidden queue of good-quality applicants, and hidden reference group comparison.

[2] Where the duration of unemployment itself is used to rank the unemployed, the wage restraint associated with any level of unemployment is reduced further. See Blanchard and Diamond (1994).

increases, but actual effort tends not to increase (Yellen and Akerlof, 1990). Although the results of the survey by Blinder and Choi (1990) should be viewed with caution,[1] it is interesting to note that while all managers agreed that workers sometimes shirk, few agreed that an increase in wages would stop shirking, but two thirds said that effort would decline if wages were reduced (Blinder and Choi, 1990).

5.3. Non-competing groups, discrimination, and unequal opportunity
According to reviewers of the efficiency wage unemployment literature:

If groups differ in their relationships between wage and productivity . . . equilibrium will be characterised by some groups being fully employed, other groups being partially employed, and still other groups being rationed out of the market. Changes in the aggregate demand for labour will have very large differential effects on the employment of different groups. (Stiglitz, 1986, p. 187).

[E]mployers may know that the functions relating effort to wages differ across groups. Then each group has its own efficiency wage and 'efficiency labour cost'; If these costs differ, it will pay firms to hire first only employees from the lowest cost group. Any unemployment that exists will be confined to labour force groups with higher costs per efficiency unit. With fluctuations in demand, these groups will bear a disproportionate burden of layoffs. (Akerlof and Yellen, 1986, p. 4).

Bulow and Summers (1986), for example, use this type of analysis to explain discrimination using their shirking model of unemployment. They suggest that because women and black workers tend to have higher rates of turnover, their time horizons are shorter, and the wage required to keep them from shirking is therefore higher than that required for other workers. In their model, the discriminatory hiring practices of primary sector employers are a rational strategy in the face of differentiated wage–productivity relationships among potential workers.

This perspective on the distribution of unemployment is limiting. In equating the incidence and duration of unemployment with being less productive at a given wage, or more expensive at a given level of productivity, efficiency wage theorists have overlooked valuable elements of dual labour market and segmented labour market theory: groups with the same actual or potential wage–productivity relationship are treated very differently in the labour market and have very different labour market opportunities and outcomes. There are at least four reasons for this: (i) the labour market is segmented by labour market internalisation so that there exist 'non-competing groups'—there is no single market which can equilibrate the treatment of all groups with the same efficiency labour cost; (ii) discrimination means that groups with the same actual or potential 'efficiency labour costs' have very different opportunities and outcomes; (iii) the supply of labour is structured and the definition of productivity and the valuation of work is not independent of the social status and the alternatives and opportunities of the group that performs it; and (iv) 'labelling' or 'scarring' can result from unemployment or from employment in secondary segments of the labour market, and can have effects on productive opportunities that are quite independent of the potential or preferences of workers.

[1] On account of small sample size—only 19 firms were surveyed.

Labour market segmentation theory would suggest that the distribution of efficiency wage related unemployment, where it occurs, has little to do with objective measures of the efficiency labour costs of different groups of workers.[1] It would suggest that the causality proposed by Bulow and Summers in their explanation for discrimination is misplaced.

6. Conclusion

Efficiency wage models of unemployment rest heavily on the original dual labour market characterisation put forward by Doeringer and Piore (1971). Although it is the dual labour market that has received most subsequent attention,[2] the fundamental part of their analysis—that the labour market is segmented by internal labour markets—is perhaps the more important idea in the context of efficiency wages. Payment of efficiency wages is one of numerous possible employer strategies in response to the recruitment, retention and motivation imperatives for labour market internalisation. Furthermore, both the internal labour market and dual labour market theories put forward by Doreinger and Piore have been developed and enhanced by subsequent literatures. Neglect of these prior and subsequent theoretical developments has clouded analysis of the incidence of unemployment.

The role ascribed to unemployment benefits in the models is also problematic. Models tend to abstract from important conditions of entitlement, and to leave the relationship between the efficiency wage and financial incentive effects of benefit levels on unemployment unspecified. From a policy point of view, the caution expressed by Dickens and Lang seems pertinent:

We have yet to derive a fully articulated and satisfactory model of labour market segmentation, and policies based on tentative models must be viewed with scepticism. Some efficiency wage models imply that it would be desirable to raise the cost to workers of being unemployed. This might be a terrible injustice if that form of the model were inappropriate. (1988, p. 133)

Bibliography

Akerlof, G. A. 1982. Labour contracts as a partial gift exchange, *Quarterly Journal of Economics*, vol. 97, 543–569

Aklerlof, G. A. 1984. Gift exchange and efficiency-wage theory: four views, *American Economic Association Papers and Proceedings*, vol. 74, no. 2, 79–83

Akerlof, G. A. and Yellen, J. L. 1986. Introduction, in Akerlof, G. A. and Yellen, J. L. (eds), *Efficiency Wage Models of the Labour Market*, Cambridge, Cambridge University Press

Akerlof, G. A. and Yellen, J. L. 1988. Fairness and unemployment, *American Economic Association Papers and Proceedings*, vol. 78, no. 2, 44–49

Atkinson, A. B. 1992. Institutional features of unemployment insurance and the working of the labour market, in Dasgupta, P., Gale, D., Hart, O. and Maskin, E. (eds), *Economic Analysis of Markets and Games*, Cambridge MA, Massachusetts Institute of Technology

[1] This is not inconsistent with the efficiency wage model of discrimination put forward by Bowles (1985).

[2] Doeringer has expressed surprise at this ascendancy: 'The concept of the internal labour market was intended to emphasise firms and unions as the principal institutions that segment the labour market, whereas the idea of the dual labour market was only an allegory for highlighting issues of poverty and discrimination' (1986, p. 48).

Atkinson, A. B. and Micklewright, J. 1991. Unemployment compensation and labour market transitions: a critical review, *Journal of Economic Literature*, vol. 29, 1679–1727

Blanchard, O. J. and Diamond, P. 1994. Ranking, unemployment duration and wages, *Review of Economic Studies*, vol. 61, 417–434

Blinder, A. S. and Choi, D. H. 1990. A shred of evidence on theories of wage stickiness, *Quarterly Journal of Economics*, vol. 105, 1003–1015

Bowles, S. 1985. The production process in a competitive economy: Walrasian, Neo-Hobbesian and Marxian models, *The American Economic Review*, vol. 75, no. 1, 16–36

Bulow, J. I. and Summers, L. H. 1986. A theory of dual labour markets with application to industrial policy, discrimination and Keynesian Unemployment, *Journal of Labour Economics*, vol. 4, no. 3, 376–414

Carmichael, H. L. 1989. Self-enforcing contracts, shirking and life cycle incentives, *Journal of Economic Perspectives*, vol. 3, no. 4, 65–83

Carmichael, H. L. 1990. Efficiency wage models of unemployment—one view, *Economic Inquiry*, vol. 28, 260–305

Dickens, W. T. and Lang, K. 1988. The reemergence of segmented labour market theory, *American Economic Review Papers and Proceedings*, vol. 78, no. 2, 129–134

Doeringer, P. B. 1986. Internal labour markets and noncompeting groups, *American Economic Association Papers and Proceedings*, vol. 76, no. 2, 48–52

Doeringer, P. B. and Piore, M. J. 1971. *Internal Labour Markets and Manpower Analysis*, Lexington, MA, Heath Lexington

Green, F. and Weisskopf, T. E. 1990. The worker discipline effect: a disaggregated analysis, *Review of Economic Statistics*, vol. 72, 241–249

Humphries, J. and Rubery, J. 1984. The reconstitution of the supply side of the labour market: the relative autonomy of social reproduction, *Cambridge Journal of Economics*, vol. 8, no. 4, 331–346

Kahneman, D., Knetsch, J. L. and Thaler, R. 1986. Fairness as a constraint on profit seeking: entitlements in the market, *The American Economic Review*, vol. 76, no. 4, 728–741

Lang, K. 1991. Persistent wage dispersion and involuntary unemployment, *Quarterly Journal of Economics*, vol. 106, 181–202

Lang, K. and Kahn, S. 1990. Efficiency wage models of unemployment: a second view, *Economic Inquiry*, vol. 28, April, 296–306

Layard, R. Nickell, S. and Jackman, R. 1991. *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford, Oxford University Press

Leibenstein, J. 1957. *Economic Backwardness and Economic Growth*, New York, Wiley

Manwaring, T. 1984. The extended internal labour market, *Cambridge Journal of Economics*, vol. 8, 161–187

Montgomery, J. D. 1991. Social networks and labour market outcomes: towards an economic analysis, *American Economic Review*, vol. 81, no. 5, 1408–1418

Okun, A. M. 1980. The invisible handshake and the inflationary process, *Challenge*, Jan–Feb, 5–13

Ramaswamy, R. and Rowthorn, R. E. 1991. Efficiency wages and wage dispersion, *Economica*, vol. 58, 501–514

Reid, G. L. 1972. Job search and the effectiveness of job-finding methods, *Industrial and Labour Relations Review*, vol. 25, no. 4, 479–495

Ryan, P. 1981. Duality and the internal labour market, in Wilkinson, F. (ed.), *The Dynamics of Labour Market Segmentation*, London, Academic Press

Salop, S. C. 1979. A model of the natural rate of unemployment, *American Economic Review*, vol. 69, 117–125

Shapiro, C. and Stiglitz, J. E. 1984. Equilibrium unemployment as a worker discipline device, *American Economic Review*, vol. 74, 433–444

Solow, R. 1979. Another possible source of wage stickiness, *Journal of Macroeconomics*, vol. 1, Winter, 79–82

Stiglitz, J. E. 1986. Theories of wage rigidity, in Butkiewicz, J. L., Koford, K. J. and Miller, J. B. (eds), *Keynes' Economic Legacy; Contemporary Economic Theories*, New York, Praeger

Wadhwani, S. B. and Wall, M. 1991. A direct test of the efficiency wage model using UK micro-data, *Oxford Economic Papers*, vol. 43, 529–548

Weiss, A. 1980. Job queues and layoffs in labour markets with flexible wages, *Journal of Political Economy*, vol. 88, 526–538

Williamson, O., Wachter, M. and Harris, J. 1975. Understanding the employment relationship: the analysis of idiosyncratic exchange, *The Bell Journal of Economics*, vol. 6, 250–278

Yellen, J. L. an Akerlof, G. A. 1990. The fair wage effort hypothesis and unemployment, *Quarterly Journal of Economics*, vol. 105, no. 2, 255–283