ISSN 0347-8769

Seminar Paper No. 503

MICROFOUNDATIONS OF UNEMPLOYMENT THEORY

by

Assar Lindbeck

Seminar Papers are preliminary material circulated to stimulate discussion and critical comment.

October 1991

Institute for International Economic Studies
S-106 91 Stockholm
Sweden

Madrid2. doc
A.L.
Oct. 25, 1991

Address to congress of
European Labor Economists,
Madrid, Sept. 26-29, 1991

MICROFOUNDATIONS OF UNEMPLOYMENT THEORY*

By Assar Lindbeck

A basic strength of the Keynesian tradition in economics is its ambition to take unemployment seriously -- both as a social problem and as an analytical issue. The notion that unemployment may be "involuntary" is one expression of this concern. But a well-known weakness of Keynesian economics has been an inability to provide a microeconomic rationale for the existence of unemployment and hence for a nonclearing labor market, with excess supply of labor and related job rationing. Associated with this weakness is the limited success of traditional Keynesian economics in clarifying in what sense unemployment is asserted to be "involuntary".

The 1980s have witnessed a resurgence of attempts to provide such microfoundations. This lecture is devoted to analyzing and extending some of these attempts. I have tried to show elsewhere how such microfoundations of unemployment theory may be integrated into macroeconomic theory; cf. Lindbeck (1992).

Before discussing rival microeconomic theories in this field, let us ask whether it is at all useful to adhere to the tradition of making a distinction between "voluntary" and "involuntary" unemployment.

* Many aspects of this lecture have been worked out over the years in cooperation with Dennis Snower; see, in particular, Lindbeck and Snower (1988).

Unfortunately, it is not an easy matter to nail down this distinction, which has led some economists to discard the concept of involuntary unemployment altogether. This illustrates, of course, the difficulty of defining "involuntary economic behavior" in general. After all, it is customary in economics to assume that the individual voluntarily, in conformity with his own preferences, does the best he can within existing constraints. This suggests, then, that the distinction between voluntary and involuntary behavior should refer to a comparison of alternative constraints for the economic agent. Indeed, such an approach forms the basis for the distinction between voluntary and involuntary unemployment in this lecture.

A useful starting point for a definition of involuntary unemployment is that it exists whenever some workers are unable to get jobs, even though they are willing to work at a lower wage than the prevailing one for jobs for which they are as equally qualified as workers who are actually employed. The unemployed workers would like to be in the shoes of the employed workers, but they are denied this possibility, through no fault of their own. This preliminary definition probably conforms to the notion among the general public that unemployment is associated with social injustice, in the sense that among people with comparable labor endowments, such as skills, some do not obtain employment on the terms applied to others.

Yet, for the purpose of the subsequent analysis, it is convenient to make this notion somewhat more precise by way of two alternative definitions.

Definition 1: At prevailing current wages and future expected wages, some workers are unsuccessful in finding jobs because, through no fault of their own, they face a more limited choice set between work and

leisure than that of employed workers, even when the wage demands of the former are adjusted for productivity differences, including unavoidable (i.e., production related) resource costs in connection with the hiring and firing of workers.

Definition 2: Some workers are unable to get jobs even though their reservation wage is lower than the value of their marginal product, adjusted for unavoidable (i.e., production related) resource costs in connection with the hiring and firing of workers.

Thus, according to the first definition, involuntary unemployment is regarded as a form of discrimination that artificially restricts the choice set of some individuals, while the second definition views it as a relation between the marginal product and the reservation wage.

Definition 2 is obviously more narrow than definition 1. If unemployed workers seek jobs at wages that are lower than their marginal product, but are unable to find such jobs, then in conformity with definition 2, these workers obviously have a more limited choice set than current jobholders with the same characteristics. This means that definition 2 is included in definition 1. On the other hand, definition 1 does not necessarily imply definition 2. For, given the unemployed workers' more limited choice set, their reservation wage may very well be higher than their marginal product, for instance as a result of the discrimination which unemployed workers may be exposed to.

In the light of these definitions, let us discuss a number of questions, in fact seven of them, that a good theory of unemployment should be able to answer:

I. Why does underbidding of prevailing wages not occur (more often) when there is unemployment? In other words, why are unemployed

workers (often) unable to get jobs by offering to work for wages below those of their employed counterparts?

II. Why are some workers often laid off rather than having their wages cut?

III. Why are wage contracts not designed to allow firms to capture all the (marginal) rents associated with employment activities, so that the wage rate is pushed down to the reservation wage for marginal workers?

IV. Why is it that unemployed workers do not "bribe" employed workers to get jobs? In other words, why do they not form coalitions with incumbent workers in order to exploit all the relevant potential gain from trade?

V. Why is unemployment not alleviated by the entry of new firms? More specifically, why are (most) unemployed workers unwilling or unable to become self-employed by starting new firms, or become employees in new firms started by others?

VI. Why do not all unemployed workers take jobs in the so-called "secondary" sector of the economy where jobs are not rationed?

VII. Why is unemployment not removed by job sharing, e.g. through job rotation or part-time work arrangements?

Let us look at how successful various microeconomic theories of unemployment are in answering these questions, among which the first deserves particular attention.

I. Why does wage underbidding not occur (more often)?

In market-clearing models, workers who tend to become unemployed have an incentive to underbid prevailing wages, and firms have an incentive to accept such underbidding, so that unemployment would

disappear. This tends to occur not only in models with perfect competition, such as in Classical and New Classical Macroeconomics and Real Business Cycle Models, but also in models that assume imperfect competition in the labor market; cf. Hart (1983), Benassy (1988), and Blanchard and Kiyotaki (1987). In all such models, it is clearly difficult to talk about "unemployment", and even more so about "involuntary unemployment" because there is no excess supply of labor and related job rationing. For instance, in models with monopolistic competition in the labor market, "syndicates" of workers are, in fact, explicitly or implicitly assumed to supply labor up to the point where the marginal revenue of the syndicate equals the marginal costs, the latter expressed by the reservation wage of the workers in the syndicate. There is clearly no excess supply of labor, and related job rationing, in this type of model, as the (identical) members of the syndicates are satisfied with the achieved combination of real wage rate and employment.

The situation is more complex in models with "implicit employment contracts", as unpredicted adverse business conditions may then lead to layoffs of some workers according to previously agreed contracts without any possibility to renegotiate. As contingent trading agreements are made before the actual transactions take place, we might say that the market clears in a special ex ante sense; given the agents' expectations, the contingent agreements are optimal for the agents. Yet ex post, when the agreements are effectuated on the basis of the economic situation that actually arises, offers to sell labor services may not turn out to be matched by offers to buy. Labor demand may fall short of labor supply, and workers who turn out to be unemployed ex post may wish to renegotiate their labor contracts, but do not (by assumption) have the opportunity to do so. Perhaps such unemployment may be called "ex ante

voluntary but ex post involuntary". It may be noted that unemployment in this case depends not merely on the emergence of unavoidable expectational errors, as in the New Classical Macroeconomics, but also on workers' assumed inability to renegotiate their wage contracts.

There are, however, theories that do explain the absence of wage underbidding and the existence of unemployment in the sense of ex ante excess supply of labor and related job rationing.

(i) In various versions of the efficiency wage theory, the general reason why workers cannot find a job is that they cannot commit themselves in a credible way to the same work effort as actually employed workers at a somewhat lower wage than the prevailing one; problems of adverse selection or moral hazard prevent such commitments. Wages may, therefore, be nonmarket clearing, with excess supply of labor and related job rationing. Unemployment in this case may be regarded as "involuntary" according to both of our definitions, at least when standard time-rate contracts are in effect. In line with definition 1, unemployed workers have a smaller choice set than employed workers; and, in line with definition 2, the reservation wage of unemployed workers may very well be lower than their potential marginal product of labor (in the hypothetical case where unemployed workers actually had the same job opportunities as currently employed workers).

An obvious strength of efficiency wage models is their appeal to common sense and everyday experience, since it is eminently reasonable to assume, as the efficiency wage theory does, that wages are used as a screening and incentive device for labor productivity. But an obvious weakness of these types of models is that they do not give employees, and their unions, a profound role in the process of wage formation. There is also the well-known question mark as to whether the efficiency wage

mechanism is quantitatively strong enough in the real world to explain much of actual unemployment -- when it is both two percent and ten percent of the labor force.

(ii) Union models imply, in fact, a tradeoff between employment and the real wage. Is it reasonable to talk about excess supply of labor and "involuntary unemployment" in this context? The answer is probably "no" in versions of union monopoly models where all workers are union members, and the union maximizes the expected utility of identical members, each with the same probability of being employed. In such models, every worker is satisfied with the combination of real wage and expected employment that has been chosen by the union. The union acts in the interest of all workers, and hence behaves as if the employee households had given the union the right to supply labor services on their behalf. This version of union monopoly models, therefore, is analytically equivalent to the model with monopolistic competition in the labor market, according to which "syndicates" of workers, rather than individual households, supply labor in the interest of individual members.

The situation is rather different in union monopoly models where individual workers have different probabilities of being employed and/or workers differ in their preferences. In this case, borrowing the terminology of Max Corden (1981), we may rather talk about "union-voluntary" but "member-involuntary" unemployment. The nature of the emerging unemployment does not change much in this case, in principle, if instead of assuming a "monopoly union" we assume that the wage rate is set by collective bargaining. However, as the wage rate in this case would normally be predicted to be lower, and the employment level higher than in the union monopoly model, the level of involuntary unemployment

would be lower. Thus, strong market power of firms in wage bargaining may be to the advantage of workers who are on the margin of becoming unemployed at a higher real wage rate.

Union models, regardless of whether we refer to union monopoly models or union bargaining models, are seriously incomplete, however, as explanations for involuntary unemployment, because they do not explain the sources of union market power. More specifically, union models do not explain why unemployed union members do not leave the union, and why nonmembers do not succeed in getting jobs by underbidding the wages of currently employed workers.

(iii) The limitations of both efficiency wage models and union models suggest the potential importance of theories that explain the sources of the apparent market power of incumbent employees; they also give unions a natural role. This is exactly what the insider-outsider theory tries to do. While efficiency wage theories build on the assumption of a specific production technology, or perhaps rather an incentive structure for the workforce, the basic assumption of the insider-outsider theory is instead that the hiring and firing of labor are associated with specific labor turnover costs. These costs give incumbent workers, "insiders", market power which they can use in their own interest to push up their wages above the potential equilibrium wage and the reservation wage of those who do not have jobs, "outsiders", without the insiders losing their jobs. The result is unemployment in the sense of excess supply of labor and related job rationing.

The most obvious type of labor turnover cost is traditional hiring and firing costs, including legally mandated severance pay to fired workers. Other, more sophisticated types of labor turnover costs may arise as a result of the process of attempted wage underbidding itself.

For instance, the firing of insiders and the hiring of outsiders may reduce work effort: a firm that is known for turning over its labor force in the event of wage underbidding may damage the work morale of its employees. Work effort may no longer be worthwhile; employees will be highly uncertain as to whether they will be able to reap the benefit of their effort.¹

Another type of labor turnover cost arises if insiders react to attempted wage underbidding by refusing to cooperate with the underbidders in the production process, as such refusals reduce the (expected) marginal product of the underbidders. Insiders may also threaten to "harass" underbidders who try to "break into" the firm by offering to work for less than the prevailing wage, and this raises the reservation wage of the underbidders.

A characteristic feature of the insider-outsider theory is that workers are analytically divided into two distinct groups for which the probability of holding a job in the future differs. Outsiders have a more narrow choice set than insiders, and the former are therefore involuntarily unemployed according to definition 1 of unemployment. We shall see later on that unemployment in some, but not all, versions of the insider-outsider model is also involuntary according to definition 2 in the sense that some workers may be unable to get jobs even though their reservation wage is lower than their labor productivity, adjusted for unavoidable (production related) resource costs in connection with worker turnover.

¹ More specifically, there is a negative substitution effect of high labor turnover on work effort, and this substitution effect may be larger than the counteracting income effect according to which a high risk of being fired means that workers cannot "afford" much shirking on the job; cf. Lindbeck and Snower, (1989, Chap. 6)

In some countries, it is illegal for firms to pay wages below those agreed on in collective bargaining, also with regard to nonorganized workers. Underbidding is then simply not legally allowed.

The notion of involuntary unemployment in the insider-outsider theory may be clarified by the following simple analogy: A little old lady is accosted by a mugger, who proposes, "Your money or your life". The lady considers her options, chooses the former, and hands over her wallet. Is this transaction voluntary? Some economists might wish to argue that it is, because the lady's action is the outcome of a "voluntary" choice. After all, she had two alternatives and she did choose one of them! But most people would probably disagree, since the lady's choice set -- through no fault of her own -- is more limited than that of other, comparable agents. The notion of involuntary unemployment in the insider-outsider theory follows the latter view. Outsiders are involuntarily unemployed (definition 1) because their choice sets are more limited than those of the insiders. In other words, there is an asymmetric social relation between incumbent workers and unemployed workers, with insiders discriminating against outsiders.

Why, then, does not the firm simply replace all its insiders by outsiders? (Reagan's firing of air controllers in the mid-1980s is one of few exceptions to the rule.) An obvious reason is that a new work force may be less productive than the one it replaces. Newcomers generally have less on-the-job training; for example, they usually have not had the opportunity to acquire the skills required to cooperate with one another in team production. On-the-job training is also often acquired by learning from insiders; but if all insiders are fired, the learning process may well be lengthy, uncertain and expensive.

Moreover, insiders' harassment activities against successful underbidders may be expected to continue even if all insiders have been fired, as fired employees are generally capable of performing such activities outside the firm as well -- at the firm's entrance or even in society at large. These activities may take the form of picket lines and social ostracism, as well as other forms of industrial and social unrest, perhaps even sabotage and violence. Attempts to replace the entire workforce may also result in a loss of customer goodwill, which may in turn be associated with a loss of firm revenues, following adverse publicity about such a replacement.

In the literature, the insider-outsider theory is frequently misunderstood to postulate that only factors inside the firm, but not outside, would influence wage setting. However, this is true solely in some special versions of the insider-outsider theory, such as the particular case elaborated by Blanchard and Summers (1986). In more general formulations, factors outside the firm also influence wage setting in insider-outsider models. For instance, the reservation wage of outsiders is influenced by conditions outside the firm, such as the unemployment rate and the system of unemployment benefits; and outsiders' reservation wage in turn affects the maximum wage that insiders can obtain without being replaced by outsiders. However, such outside conditions also influence the income of insiders in the event that they would lose their jobs. In a stochastic framework, this will affect the aggressiveness with which insiders, in their own interests, are willing to push for higher wages. All this means that the insider-outsider theory should not be regarded as pure "insider theory", but literally an "insider-outsider theory".

These misunderstandings have also distorted some empirical tests of the theory. This has occurred when it has been asserted that the theory would imply that factors outside the firm do not influence wage setting. It has also occurred when the tests have assumed that the insider-outsider theory necessarily predicts that the wage rate is a negative function of the insider workforce in the preceeding period.

As a criticism of the insider-outsider theory, it has also been argued that threats of noncooperation and harassment against underbidders may not be time consistent and therefore not credible, as noncooperation and harassment may be disagreeable forms of behavior even for those who take part in them; cf. Fehr (1990). In reality, however, there are reasons to assume that noncooperation and harassment strategies are indeed time consistent, as insiders do have something to gain from implementing earlier announced threats of noncooperation and harassment after a firm has actually employed an underbidding outsider. First, it is important to note that, in general, even if noncooperation and harassment are probably disagreeable activities for the harasser himself, this may not always be the case. Indeed, there may be situations where an insider could even derive a positive utility from harassment, i.e., precisely when he is provoked by underbidding entrants. Workers who find that their jobs or wages are threatened by underbidders, often called "scabs" or similarly derogatory terms, may regard it as a relief to harm underbidders by noncooperation and harassment. In this event, threats of noncooperation and harassment may very well be credible.

Furthermore, even if noncooperation and harassment activities would always be disagreeable to the insiders, threats of such actions may nevertheless be credible in a multiperiod setting. In such a context which, of course, is more realistic than a one-period setting, the

insider may have a current incentive to implement noncooperation and harassment threats in order to establish a reputation in the future. In particular, as is easily shown, insiders will have an incentive not to cooperate with, and to harass, entrants whenever the disutility from doing so is less than the utility from the associated rise in the future insider wage; see Lindbeck and Snower (1990).

(iv) All the microtheories of unemployment discussed above rely on the notion of rational economic optimization. However, it is important to realize that a great deal of human behavior is based on habits and social norms. This aspect has indeed been considered in some versions of both the efficiency wage theory and the insider-outsider theory. In George Akerlof's (1982) version of the efficiency wage theory, it is assumed that the efficiency of work is largely a result of social attitudes, and social norms among employees, and that these norms are based on an exchange of "gifts" between firms and their employees. Firms "give away" wage payments in addition to the reservation wage of workers and the potential market equilibrium wage, while workers give the firm a pro quo gift in exchange in the form of extra work effort.

Whereas Akerlof's version of the efficiency wage theory is built on assumed norms among incumbent workers concerning wages and work effort, the noncooperation/harassment version of the insider-outsider theory instead establishes wage norms among unemployed workers. Thus, rather than merely assuming that such norms exist, the insider-outsider theory tries to explain the sources of these norms; i.e., it tries to establish in whose interest the norms are created and how they are upheld. According to the noncooperation/harassment version of the insider-outsider theory, norms are created to protect the interest of insiders

vis-à-vis outsiders; and these norms are upheld by insiders' threats of noncooperation and harassment toward outsiders who try to "break into" firms by offering to work for a lower wage than the current insider wage.

It has been shown by Jörgen W Weibull (1987) and Robert Solow (1991, pp. 38-50) that a Nash equilibrium may emerge in which unemployed workers abstain from underbidding in the expectation that they themselves may then, if employed in the future, not be threatened by wage underbidding from others. However the Weibull-Solow equilibrium does not really explain the dynamic incentive mechanisms that make people abstain from underbidding: an unemployed individual hardly has an incentive to abstain from the personal gains associated with successful underbidding on the grounds that his own behavior (in a world with many workers) would create incentives for other individuals to act in a similar fashion in the future (when he may have a job).

The problem is rather similar to the "voting paradox", i.e., the difficulty in explaining why an individual chooses to vote even though the probability that he will influence the election outcome is negligible. However, while we may argue that the cost of voting is also negligible, and even that people may actually enjoy the act of voting, we can hardly argue that the costs of not underbidding are negligible. Using another metaphor, it may be claimed that people shovel snow away from their sidewalk in the expectation that their neighbors will do likewise. Here, however, we have an example of social control within small groups, in contrast to the hypothesis that an unemployed worker would abstain from wage underbidding in the expectation that others would not underbid his wage when he has a job in the future.

It would therefore seem that more convincing micro-based disincentives to underbidding behavior have to be identified and

explained. The noncooperation and harassment version of the insider-outsider theory may do just that.

II. Why are some workers laid off?

A successful explanation of involuntary unemployment should not only be able to explain the existence of unemployment. It should also tell us why lay-off of workers is not avoided through wage cuts. Both efficiency wage theories and insider-outsider theories do provide such explanations. According to the efficiency wage theory, the reason is that such wage cuts would be unproductive for firms if the fall in the average productivity of the workforce initiated by the wage cut is sufficiently large as compared to the fall in the average wage rate. According to the insider-outsider theory, remaining (i.e., nonfired) incumbent workers may prevent wage underbidding by refusing to cooperate with, or by threatening to harass underbidding fired workers -- in the same way as insiders may exercise such powers against underbidding attempts by initially unemployed workers. This means that we may talk about different categories of insiders, where senior workers are characterized by more "insidedness" than junior workers. Thus, the explanation of layoffs is basically the same as the reason for the existence of unemployment in general -- in both efficiency wage theories and insider-outsider theories.

A complication inherent in the insider-outsider theory explanation of lay-offs should also be mentioned. If senior insiders are anxious to keep their wages unchanged in a business downturn, even though this results in layoffs of junior insiders, why do senior insiders refrain from pushing up their wages long before, as they are assumed not to care much about fired workers anyway? There are several potentially

reasonable answers to this question (which, by the way, resembles the notorious question of the "shrinking" union with median voter decisions and seniority). One conceivable answer is that if insiders succeed in pushing up their wages above a certain level, defined by the sum of the entrant wage of outsiders and the marginal hiring and firing costs, it will be profitable for the firm to replace its insiders by outsiders. Moreover, it may be the case that there are limits to the egotism of senior insiders. They may be egotistic enough to prevent underbidding attempts by fired insiders, but not egotistic enough to take the initiative in pushing up wages to the extent that junior insiders lose their jobs. For instance, noncooperation and harassment activities may be a defensive action against underbidding, but not an offensive action to reduce the labor force by aggressively pushing up the wage rate.

III. Why are contracts not designed, to allow firms to capture all the (marginal) rents associated with employment activities?

If firms could capture all the (marginal) rents from employment, marginal workers would become indifferent between work and leisure at a wage rate that corresponds to the marginal product of labor, and involuntary unemployment according to definition 2 would disappear. There has been much discussion in the literature as to whether unemployment in this sense could be avoided through more or less complex contracts between firms and their employees in the context of either efficiency wage or insider-outsider models.

One example may be long-term wage contracts in which an entrant agrees to a particular wage trajectory covering his entire period of employment in a firm. The contract would be such that each entrant is just as well off with work throughout the period in question as he would

be without work. The idea is that regardless of what wages senior workers receive, there is always, in principle, an entrant wage that is sufficiently low to make the entrant indifferent between work and leisure over the entire span of the prospective employment period, so that the firm captures all the rents from employing the worker. In the case of perfect capital markets, only the present value of the wage trajectory is relevant for the utility of the worker, as well as for the labor cost of the firm. As a result, it may be argued, the distribution of the complete package among various years is unimportant.

If this entrant wage happens to be negative, which is quite likely, we have the familiar case of "entry fees", whereby entrants pay their firms a particular sum of money for the privilege of receiving employment. In a variation on this theme, so-called "bonding", an entrant hands over a deposit to the firm when he is hired -- a deposit which is foregone if he leaves the firm prior to retirement (for reasons other than physical incapacity). The firm would set the fee, or deposit, sufficiently high so that an entrant becomes indifferent between the prospect of employment and unemployment. Real world counterparts, or rather substitutes, for such contracts are various types of "postponed" wage income, including pension rights that are lost if the individual worker quits the firm.

By pointing out the possibility of such contracts and remuneration schemes, it is sometimes asserted that efficiency wage and insider-outsider theories do not really explain unemployment (of type 2), but rather the intertemporal wage profile of workers. A related criticism, referring to unemployment of type 1, is that outsiders are not really involuntarily unemployed because they may obtain the same expected flow of lifetime utility as incumbent workers, provided that the utility flows

are evaluated at the beginning of each agent's respective working lifetime. The argument is that every incumbent was once an entrant, and if all entrants receive their reservation wages, they are by definition indifferent in the choice between work and leisure over their lifetime. Consequently, by this argument, entrants would not really be discriminated against.

To elaborate on these points, suppose that insiders suddenly find new rent-creating opportunities and consequently drive up the insider wage. As result, in an intertemporal setting, the reservation wage of entrants falls because the higher the wage that workers expect to receive in the future, the lower the wage they are willing to accept at present. Now suppose that entrants receive exactly this low reservation wage. Then the firm's present value of expected wage payments over the rest of a worker's lifetime may remain unchanged. The higher wage that the worker receives as an insider would be fully counterbalanced by a lower wage as an entrant. In this event, the firm's employment may also be unchanged. If so, the insiders' influence over their wages would not create unemployment, but merely steepen the intertemporal wage profile.

To what extent are efficiency wage models and insider-outsider models open to such criticisms? A kind of "minimum defence" of the efficiency wage and insider-outsider theories against this criticism would be that these theories simply try to explain the existence of involuntary unemployment in a world of standard time contracts -- without entrance fees or bonding -- and that real world contracts with deferred payments (as argued by Katz and Akerlof, 1990) are not perfect substitutes for entrant fees and "bonding". Another, and perhaps more profound, answer involves showing that contracts like those suggested above are seldom easy to agree on and implement due to problems of moral

hazard, time inconsistency and monitoring. Moreover, in many countries, particularly in Western Europe, contracts where pensions are lost when employees leave their firms may not be allowed legally. Furthermore it has often been pointed out that credit markets may not be "perfect", so that the individual might not be able to borrow the sum of money to pay a requested entrant fee.

Moreover, what is supposed to be captured by the insider-outsider theory is the notion of inequality of opportunity over a period starting at a particular time, when some workers have jobs and others do not. It should be kept in mind that the definition of unemployment in this theory involves a comparison of the wage-labor service opportunities of an outsider and a current insider over a unique set of time periods. It is not concerned with a comparison of these opportunities over an outsider's future working life and those over an insider's past, present and future working lifetime. I believe that everyday notions of unemployment are more readily captured by the former comparison than by the latter.

It may also be argued that the present value of a firm's wage costs over a worker's remaining lifetime in the real world is positively related to the insider wage, which is another way of saying that the entrant wage does not fully adjust downward to rising insider wages. There are several reasons for this. For example, workers may have a higher rate of time discount than firms, so that a rise in the insider wage may not reduce the reservation wage sufficiently to keep the firm's present value of wage costs (under the firm's rate of time discount) from falling. Perhaps more important, the entrant wage may exceed the intertemporarily calculated reservation wage, and perhaps even coincide with the insider wage, because the insiders use their market power to exert upward pressure on the entrant wage in order to discourage the

entry of low-wage workers. As a result, firms may be unable to provide a remuneration package that makes any of its workers indifferent between work and leisure. This is particularly likely if insiders are supported by unions which can close down the firm by strikes, or severely damage production by work-to-rule, or stop deliveries to the firm from other firms by union-organized blockades. In all these cases, as well as in the previously discussed case with a "liquidity constraint", the insider-outsider theory certainly does explain unemployment (according to definitions 1 and 2), and not just the intertemporal wage profile.

IV. Why is it that involuntarily unemployed workers do not "bribe" incumbent workers to get jobs?

As unemployment, in terms of the definitions above, is a Pareto-inefficient state, we might think that unemployed workers could promise incumbent workers a share of the income gains that they would receive if they became employed. The question then is why the agents in market economies do not take advantage of such opportunities. Where does the distortion lie that makes the private benefit from formation of such a coalition smaller than the social benefits?

One explanation may be a situation where incumbent workers are afraid that in the future, when the number of low-wage workers has become relatively large, the firm will replace the remaining high-wage workers by low-wage workers. Thus, permanent two-tier wage contracts may not be accepted by incumbent workers. If, on the other hand, incumbent workers would accept such contracts, the low-wage workers might refuse to accept the two-tier wage system in the future, as their market powers grow when they gradually achieve insider status in terms of equally high labor turnover costs as those of the initially incumbent workers. For this

reason, firms may not be interested in permanent two-tier remuneration systems.

Still, we may be able to conceive of more or less ingenious wage contracts that could induce insiders to cooperate fully with, and forego harassing underbidding entrants, thereby eliminating the involuntary unemployment in the noncooperation and harassment version of the insider-outsider model. Indeed such contracts have been suggested in the literature; see Fehr (1990).

One such contract would imply that firms promise to give an additional wage premium to workers who cooperate fully with, and do not harass underbidding entrants. However, cooperative activities are the result of teamwork among workers, and it is an inherently impossible task to identify an individual's contribution to a cooperative effort with sufficient objectivity and accuracy to use that observation as a basis for wage payments. Thus, the problem is not only that there is uncertainty as to whether a noncooperator will be caught or not; it is usually also impossible to prove that a worker who is assertedly caught for shirking has in fact been shirking. The difficulties involved in observing harassment activities are, if anything, even greater since the final outcome of these activities (a rise in the entrant's disutility of work) is not open to objective measurement.

It is worth noting, however, that whereas the inputs into team production cannot be measured objectively in practice, the output may be measurable in some cases. Output-related wage contracts therefore appear to be more promising devices for inducing insiders to cooperate with entrants. Yet such contracts also give rise to serious problems of implementation. Output-related wage contracts are never available to reduce unemployment generated through harassment activities, as the firm

is unable to infer the existence of such activities from output variables. The matter is not quite so clear-cut for cooperation activities. Although the firm cannot observe these activities directly, it is able to observe its total output. Thus, it may be able to reward its insiders for cooperating with entrants by sharing the proceeds of its output with these insiders. For instance, this could take the form of profit or revenue sharing. Yet, there are a variety of obstacles to designing and implementing such output-related wage contracts. Consider three significant impediments.

First, there are monitoring difficulties. Since profit and revenue sharing schemes are generally difficult and costly for workers to monitor, managers may have an incentive to use their superior position in computing profit or revenue figures to their own advantage. In response, the employees may have an incentive to implement monitoring procedures themselves (and possibly also to engage in litigation). The gains from profit sharing may not fully compensate the firm and its employees for these monitoring costs. The managers of firms may also be unwilling to let workers and their union have the degree of insight into the operations and bookkeeping of the firm that is required to solve these monitoring problems.

Second, profit and revenue sharing schemes inevitably also involve the imposition of additional risks for employees. If the employees are risk averse, then they suffer a utility loss in connection with increased uncertainty regarding their remuneration. The firm may be unable to compensate them for this loss without robbing itself of the potential gains of such schemes.

Third, such schemes also have consequences for the market power of insiders. When an insider decides to cooperate with entrants, he loses

something and gains something. He gives up market power vis-à-vis the entrants, although he gains some of the profit or revenue that accrues as a result of his cooperation with entrants. The second effect outweighs the first only if the firm relinquishes at least a certain amount of its additional gross profits. Yet if the firm does so, it may find that its net profit is lower than in the noncooperative equilibrium, and then it has no incentive to implement the contract.

Difficulties like these help explain why output-related wage contracts do not play a particularly prominent role in today's labor market. However, there is no reason to believe that the difficulties are always insuperable; indeed the model of differential cooperation activities suggests that there is a real-world case to be made for seeking alternatives to time-rate contracts, in particular productivity- or profit-related contracts.

V. Why are involuntarily unemployed workers not absorbed by new firms?

If workers were always able to become self-employed, then there could be no involuntary unemployment. Whoever could not work for someone else could work for himself. In principle, it is probably always possible to find some form of self-employment. As extreme examples, most people have the option of becoming street peddlers or flower sellers of from an improvised stand somewhere. Yet a nuclear physicist, a school teacher or a skilled worker who, after having been laid off, prefers leisure to selling flowers or newspapers on the street, would probably still, by most people, be classified as involuntarily unemployed.

The question remains, however, as to why workers are often unable or unwilling to start new firms, and why new firms started by others do not employ the unemployed workers. Weitzman (1987) has argued that

increasing returns for firms operating under imperfect competition, in combination with insufficient domestic aggregate demand, may prevent new firms from entering. However, considering the enormous heterogeneity of the production system, products and the structure of demand, it sounds rather implausible that returns to scale dominate for most types of products. In addition if enough firms happen to be started simultaneously in the national economy, the incomes that these firms generate should, in principle, also create aggregate purchasing power and aggregate demand to sustain many of these firms, even if it would be extreme to assume that Say's law holds fully. Moreover, in an open economy, there is the possibility that new firms can replace imports, or even engage in exports, which would provide markets without concern for domestic aggregate demands.

It is, for these reasons, important to find other explanations for why entry of firms does not remove unemployment completely. One simple, but perhaps realistic answer is that the process of starting and running firms requires specific organizational abilities and a greater willingness to take risks than many workers have. Difficulties in acquiring equity capital and loans comprise another explanation. Moreover, legislation or labor union activities in some countries may force new firms to pay the same wages as in already existing firms to unorganized workers, which reduces new firms' possibilities of entering. Most likely, there are other important obstacles to the entry of firms, for instance in connection with the tax system and various government regulations. For the time being, however, we have rather rudimentary knowledge about the importance of different types of barriers to the entry of firms; moreover, as the obstacles are partly institutional in nature, they would be expected to differ among countries.

Nevertheless, considering that the entry of new firms, where by definition there are no insiders, is potentially important for aggregate employment, policies which promote employment have good reason to try to remove obstacles to the entry of firms, regardless of what these obstacles might be.

VI. Why do involuntarily unemployed workers refrain from taking jobs as employees in the "secondary" sector?

Union theories, efficiency wage theories and insider-outsider theories are designed to explain why labor demand may fall short of labor supply in those parts of a national economy where unions are important, or where efficiency wage or insider-outsider mechanisms are important aspects. Hence these theories are designed to explain why jobs are rationed in the so-called "primary sector" of the economy. By themselves, however, these theories do not explain why some workers choose to remain unemployed rather than work in the "secondary" sector of the economy, i.e. in firms where unionization is insignificant, efficiency wage effects are weak, and labor turnover costs, and hence insider market power, are small. These theories therefore have to be amended to explain how workers, who are shut off from the "primary" sector, are distributed between open unemployment and work in the secondary sector, including self-employment in the latter.

One way to explain this is to combine some of the theories just discussed with the Harris-Todaro model of labor market segmentation. Such a combination has in fact been proposed in the case of efficiency wage models; cf. Bulow and Summers (1986). However, the Harris-Todaro theory builds on the rather implausible assumption that unemployed workers are distinctly better able to search for primary-sector jobs than

are workers employed in the secondary sector. Therefore, perhaps labor market segmentation should be explained in some other way. A particularly simple approach is to use a model where outsiders are "sorted" into open unemployment and work in the secondary sector by workers' heterogeneity in terms of taste, productivity and wealth. This issue is illustrated in Figure 1, where LD_p and LD_s are the demand functions in the primary and secondary sectors, WS is the wage setting function in the primary sector and LS_s is the labor supply function in the secondary sector. The entire horizontal distance OO' denotes the size of the entire labor force. The model then determines employment in the primary sector as OA and employment in the secondary sector as $O'B$, with AB denoting the size of unemployment.

Figure 1

This is not a very deep explanation. It becomes somewhat more profound if it can be argued that the heterogeneity of workers is endogenous with respect to the previous employment and unemployment experience of workers; cf. Lindbeck and Snower (1988). For instance, those who used to work in the primary sector may have developed a particularly strong reluctance to work in the secondary sector, while those who have worked in the latter sector may have developed less aversion for such work, and also acquired sector-specific skills that boost their productivity for secondary sector work. Moreover, workers previously employed in the primary sector may have acquired more wealth than those who have worked in the secondary sector, which may make the former choose differently than the latter, e.g. as they are more able to finance open unemployment, or work in the household sector, rather than work in the secondary sector (a wealth or liquidity effect).

Figure 1: Dual labor market

In some countries, strictly administrated minimum wages or legislation that makes collective agreements binding also for the secondary sector may help explain why some workers wind up as openly unemployed rather than as employed in the secondary sector. In this specific case, however, there would be excess supply and job rationing not only in the primary but also in the secondary sector.

VII. Why is involuntary unemployment not reduced by job sharing, e.g. through job rotation or part-time work arrangements?

Although the work week is sometimes shortened when business conditions turn adverse, layoffs and dismissals are nevertheless common. Besides, firms rarely rotate individuals between employment and unemployment. What is the reason for this? In other words, why are workers laid off, rather than having the number of working hours cut for everybody.

If workers are risk averse, they would generally prefer job sharing (and thereby receive a comparatively steady income flow) to the prospect of being employed during some periods and unemployed during others. Moreover, job-sharing would benefit the outsiders, not only because they may obtain a higher income by part-time work than by being unemployed, but also because as part-time employees they can acquire or keep their status as insiders, and hence be counted among the privileged group of workers in wage formation and employment decisions in the future. Why then do firms not satisfy the likely preference for part-time work among laid-off workers?

The various microeconomic unemployment theories discussed above have hardly dealt with this issue. A common answer in the labor market literature is that fixed costs in connection with work make it more

costly for firms, or workers, to cut the number of hours worked by individual employees than to reduce the number of employees. Another and perhaps more important explanation, which follows directly from the insider-outsider theory, is that senior insiders, who feel that their jobs are safe, may use their market power to prevent part-time work from expanding in a business downturn. Thus, a reason why part-time work is not more prevalent than it actually is during recessions may, again, be a conflict of interest between employees with different degrees of seniority, i.e. with different degrees of "insidedness".

VIII. Concluding Remarks

The punchline of this address is that nowadays there are promising microeconomic theories with the ambition of explaining unemployment, in the sense of excess supply of labor and related job rationing. I have argued that the efficiency wage theory and the insider-outsider theory are the most promising examples of such theories. These theories also explain why workers are often laid-off or fired, rather than retained at lower wages than the initial one when a reduction in the demand for labor occurs. Moreover, the type of unemployment that is explained by these theories may be called "involuntary" in a well defined way.

Although neither the efficiency wage theory nor the insider-outsider theory requires unions, the insider-outsider theory turns out to provide an analytical underpinning of union models, by explaining where union power may come from, i.e., from labor turnover costs. By way of political lobbying for improved job security legislation, unions may raise labor turnover costs. Moreover, they may organize noncooperation and harassment activities by insiders against underbidding outsiders more efficiently than can be achieved without unions, as well as provide new

tools for seizing a larger share of firms' rents, for instance by threats of strikes or work-to-rule. This implies that the insider-outsider theory may also be regarded as an attempted contribution to institutional economics, in the sense that it may help explain the emergence of labor unions. It is, finally, also possible to combine efficiency wage theories, insider-outsider theories and union theories in a unified analytical framework; cf. Lindbeck and Snower (1991).

Needless to say, all these theories may be improved considerably. It may also be important to integrate them with search theory. Moreover, much work remains. In particular they have to be confronted with empirical data in relevant ways which, unfortunately, is by no means an easy task.

References

- Akerlof, G. (1982) "Labor Contracts as Partial Gift Exchange", Quarterly Journal of Economics, 97, pp. 543-569.
- Benassy, J.-P. (1988), The Economics of Market Disequilibrium, New York: Academic Press.
- Blanchard, O.J. and Kiyotaki, N. (1987) "Monopolistic Competition and the Effects of Aggregate Demand", American Economic Review, 77(4), pp. 647-666.
- Blanchard, O.J. and Summers, L. (1986), "Hysteresis and the European Unemployment Problem", NBER Macroeconomic Annual, Vol. 1, MIT Press: Cambridge MA.
- Bulow, J. and Summers, L. (1986), "A Theory of Dual Labor Markets with Application to Industrial Policy Discrimination and Keynesian Unemployment", Journal of Labor Economics, 4, pp. 376-414.
- Corden, M. (1981) "Taxation, Real Wage Rigidity and Employment", Economic Journal, 91, pp. 309-330.
- Fehr, E. (1990) "Cooperation, Harassment, and Involuntary Unemployment: Comment", American Economic Review, 80(3), pp. 624-30.
- Hart, O.D. (1982) "A Model of Imperfect Competition with Keynesian Features", Quarterly Journal of Economics, 97, pp. 109-138.
- Katz, L. and Akerlof, G. (1989) "Workers' Trust Funds and the Logic of Wage Profiles", Quarterly Journal of Economics, Aug. pp. 525-536.
- Lindbeck, A. (1992) "Macroeconomic Theory and the Labor Market", Presidential Address EEA, European Economic Review, Papers and Proceedings (forthcoming).
- Lindbeck, A. and Snower, D. (1988), The Insider-Outsider Theory of Employment and Unemployment, Cambridge, Mass: The MIT Press
- (1990) "Cooperation, Harassment, and Involuntary Unemployment: Reply", American Economic Review 80(3), pp. 831-836.
- (1988) "Segmented Labor Markets and Unemployment", Seminar Paper No 483, Institute for International Economic Studies, University of Stockholm, November.
- , (1991), "Interaction Between the Efficiency Wage and the Insider-Outsider Theories", Economic Letters.
- Solow, R. (1990), The Labor Market as a Social Institution, Oxford: Basil Blackwell.

Weibull, J. W. (1987) "Persistent Unemployment as Subgame Perfect Equilibrium", Seminar Paper No 381, Institute for International Economic Studies, Stockholm , May.

Weitzman, M. (1987), "Steady State Unemployment under Profit Sharing", Economic Journal, 97, pp. 86-105.