Insiders versus Outsiders

Assar Lindbeck and Dennis J. Snower

Abstract

This article surveys the insider-outsider theory, which analyzes the behavior of economic agents in markets where some participants have more privileged positions than others. Incumbent workers (insiders) in the labor market enjoy more favorable employment opportunities than others (outsiders), on account of labor turnover costs (e.g., costs associated with hiring, training, firing, and insiders' ability to punish underbidding outsiders). The theory provides an explanation of existence and persistence of unemployment, and helps account for the different employment experiences in Europe and the United States. The article also discusses policy implications and examines how the theory is related to unions and social norms.
The insider-outsider theory examines the behavior of economic agents in markets where some participants have more privileged positions than others. In the labor market, incumbent workers, the "insiders," often enjoy more favorable employment opportunities than the "outsiders." The reason for this disparity is costs that firms incur when they replace insiders by outsiders – labor turnover costs for short. These can come in many forms, the most obvious ones being the costs of hiring, firing and firm-specific training, but they can also arise from insiders’ attempts to resist competition with outsiders by refusing to cooperate with or harassing outsiders who try to underbid the wages of incumbent workers. Since - for reasons to be explained below - these costs are borne, at least in part, by the employers, they give the insiders market power. The insiders use this power to push their wage above the market-clearing level, but firms do not try to replace them with outsiders since it would be costly to do so. The insider-outsider theory then proceeds to examine the implications of this behavior for employment and unemployment (Lindbeck and Snower, 1984, 1986).

The theory was originally constructed as a microeconomic foundation of the existence of unemployment, hence an explanation for the absence of wage underbidding even though many unemployed workers would like to work for wages lower than existing insider wages, normalized for productivity differences. The theory, however, has also been applied empirically to more specific labor market issues, like the different employment experiences in Europe and the United States in
recent decades.

The insider-outsider distinction refers to a wide number of divides: employed versus unemployed workers, formal- versus informal-sector employees, employees with high versus low seniority, unionized versus non-unionized workers, workers on permanent versus temporary contracts, skilled versus unskilled workers, the short-term versus the long-term unemployed, and so on.

These distinctions also translate into social differences. In many developed countries nowadays growing attention is devoted to the phenomenon of “social exclusion”. Some individuals, families and other social groups are excluded from the mainstream networks of social relations within a society. They are typically unemployed or working at temporary, low-grade, or dead-end jobs, and finance much of their consumption out of transfer payments (e.g. from social assistance programs, their parents’ incomes) the black market, or even criminal activities. They often live in the underclass neighborhoods of large cities, with meager social services, poor schooling, and scant police protection. These are the real “outsiders” in society, and their outsider position in the labor market is an important source of their social exclusion. These consequences are sometimes accentuated by rent control, which creates insiders and outsiders in the market for rented apartments as well; this group comprises individuals with poor networks, often as a result of their weak labor market position.

In labor economics there are three main ways of explaining why some workers face less
favorable employment conditions than others do. These correspond to the three general reasons why firms may be unwilling to replace their current employees by outsiders, even when the outsiders are willing to work at wages that would more than compensate the firm for insider-outsider productivity differences: First, legislation may keep the wage above its market-clearing level (the minimum-wage explanation). Second, firms may not accept the outsiders’ underbidding, since a fall in the wage may reduce productivity or increase the rate of labor turnover (the efficiency wage explanation). And third, it is not in the insiders’ interests to permit outsider underbidding and the insiders can impose their interests on their employers, since the insiders’ positions are protected by labor turnover costs (the insider-outsider theory explanation). The labor union and wage bargaining literature falls into the third category of explanation, since the insider-outsider theory provides a rationale for unions, explains what gives unions their clout and identifies sources of unions’ wage bargaining power.

But the theory applies to far more than the labor market and the housing market. In general, whenever a buyer faces a cost of switching from its current supplier to another one, the current supplier thereby gains market power, which can be exploited by raising the price. For example, if product or credit market regulations make it costly for new firms to enter, then consumers will find it costly to abandon the existing suppliers in favor of new suppliers, and thereby the existing suppliers gain market power. Just as product markets become less than perfectly contestable when there are barriers to the entry of firms (Baumol, Panzar and Willig (1982)), so labor markets become less than perfectly contestable when there are labor turnover costs.
This article is an idiosyncratic survey of the insider-outsider theory. Its overriding aim is not to provide a comprehensive summary of the relevant literature, but rather to describe the vision underlying the theory and to discuss salient contributions to the literature in the light of this vision. The rest of this article is divided into four sections. The first concerns the theory of how labor turnover costs influence wages, employment and unemployment. The second section deals with the insider-outsider theory in relation to two important economic institutions: unions and social norms. The third section confronts the relevant empirical evidence. The last section concludes by briefly presenting some policy implications.

**The Theory**

Insider-outsider models rest on the following central building blocks:

1. *Types of contract*: When an outsider is hired, he becomes an entrant. Once an entrant has remained with the firm for a period sufficiently long to become associated with the same labor turnover costs as the insiders, the entrant is assumed to have the opportunity to renegotiate the wage. In effect, the entrant thereby turns into an insider. Periodic renegotiation of wages arises partly for legal reasons: employment law usually allows for renegotiation of contracts (see, for example, Malcomson (1997)). Another reason is that it is usually impossible in practice to write contracts contingent on all the possible future events that may be relevant to the employment relationship. In practice, of course, the distinction
between insiders and entrants is a matter of degree, with gradations of insider power reflecting the gradual rise of labor turnover costs with job tenure. Insiders who move from firm to firm may retain some of their insider status, depending on the specificity of their human capital, the provisions of job security legislation, the coverage of union agreements, and so forth.

2. Production- versus rent-related turnover costs: The turnover costs may be divided into “production-related” costs, which must be expended in order to make outsiders productive within a firm, and “rent-related” costs, which are the outcome of insiders’ rent-seeking activities. Both are common in most labor markets (see Malcomson (1998) for evidence).

3. Incidence of labor turnover costs: Firms generally cannot pass all labor turnover costs on to their insiders. One reason is that the firms do not incur these costs until they replace their insiders with newly recruited entrants. Furthermore, whereas firms could, in principle, shift labor turnover costs to their insiders through insider “exit fees” – lump sum payments by insiders to their employers upon voluntary quitting – such fees are usually illegal and incentive incompatible. Incentive incompatibility arises because it is often difficult to assess whether a separation is generated by voluntary quitting or firing “without cause”, and consequently exit fees would encourage firms to replace their insiders by entrants in order to obtain such fees.

4. Market power: Since firms bear some of the labor turnover costs, insiders’ jobs are protected by these costs. Since entrants’ jobs are less protected, insiders have more market power than
entrants. Insiders use their market power to drive their wages up.

5. **Employment decisions:** These decisions are made unilaterally by firms.

On account of the rent-related turnover costs, the outsiders face labor market discrimination in the sense that they do not receive equal reward for equal productivity. As a result, outsiders may be involuntarily unemployed or involuntarily confined to dead-end jobs, i.e. they may be unable to work their way into insiders’ jobs even though they are willing to work for less than insider wages, normalized for productivity differences and production-related turnover costs. The outsiders are not party to the negotiations that set insiders’ conditions of employment. The insiders, by contrast, can engage in various anti-competitive practices – e.g. claiming more than the market-clearing wages, pushing for seniority rules, or driving entrants wages above the minimum levels necessary to attract them – without necessarily facing dismissal.

Let us take a closer look at how labor turnover costs lead to such inequalities.

**Labor Turnover Costs and Insider Wage Formation**

Whereas the literature on human capital acquisition, search, and labor market adjustment primarily considers production-related labor turnover costs, such as search, hiring, relocation and training costs, the insider-outsider theory also focuses on rent-related costs. The latter may come in many forms, such as severance pay, seniority rules, requirements that firms give insiders
advance notice of dismissal, and other forms of legal protection against firing. Insiders, who generally comprise the majority of a developed country’s electorate, often use their political influence to raise such costs (Lindbeck and Snower, 1988b, ch. 11; Saint-Paul, 1996).

Beyond that, insiders can often protect their jobs and prevent wage underbidding by cooperating with one another in the production process (thereby raising each others’ productivity), but refusing to cooperate with outsiders who offer to work for less, creating an insider-entrant productivity differential. Insiders may also be on friendly terms with one another but “harass” underbidding new recruits, thereby generating an insider-entrant reservation wage differential that translates into a labor turnover cost (Lindbeck and Snower (1988a)). Labor turnover costs associated with non-cooperation and harassment activities are likely to be particularly significant whenever work is performed in teams; these costs are manipulable by the insiders and do not require the existence of unions or job security legislation. Clearly, significant rent-related labor turnover costs exist also in the United States, even though job-security legislation and unions are relatively weak, and hence insiders have less market power than in Europe.

To see why labor turnover costs are the ultimate source of insiders' market power, it is useful to recognize that a firm has two alternative partners in wage negotiations: insiders and outsiders. Labor turnover costs determine the degree of substitutability between these two alternative negotiations. When turnover costs are zero, the two sets of negotiations are perfect substitutes and
the insider wage is driven down to the reservation wage of the marginal worker. When these costs are prohibitively high, the firm and its insiders have bilateral monopoly power. In between these two extremes, the market power of insiders may be regarded as rising with labor turnover costs (Manzini and Snower, 1998).

The Influence of Insider Power on the Outsiders' Opportunities

When insiders use their market power to improve their labor market opportunities, what happens to the opportunities of the outsiders and how do entrant wages respond?

If the entrants had no market power at all and there were no lower bounds to their pay, the marginal entrants would always receive their reservation wage, making them indifferent between staying jobless over their remaining lifetime and the expected present value of remuneration from work. Then insiders' power over their wages would have no influence on the present value of labor costs for entrants; any rise of the insider wage would simply be offset by a fall in the entrant wage. Thus firms’ employment decisions would be unaffected. This point has been made by Bertola (1990) and others.

However, if firms bear some of the labor turnover costs, they cannot hand all these costs on to the entrants. Moreover, since insiders' wages are usually well above the amount necessary to keep them from dropping out of the labor force, and since these wages are paid over much of their working lifetimes, an entrants' reservation wage (before becoming an insider) is usually very low –
in general substantially negative. But in practice we do not observe entrants paying large sums of money to purchase jobs from their employers. There are various reasons why this is so.

First, entrants may be unable to afford a large, negative reservation wage on account of credit constraints. Second, such a reservation wage may be infeasible on account of minimum wage laws, social norms, or union pressure on firms. Third, entrants may have some market power, since firms often expend costs in hiring and training them (e.g., advertising, screening, and on-the-job training costs). Indeed, if these costs are incurred before the firm negotiates the wage with the entrant then, by the time the wage negotiations occur, these costs are sunk and therefore the firm cannot shift them onto the wage, according to standard bargaining theory.³

Fourth, firms may offer entrants more than their reservation wages to avoid an incentive incompatibility problem (Manzini and Snower, 1998). Entrants will have one reservation wage if they anticipate that they will become insiders - thus remaining "bonded" to the firm and receiving insider wages in the future. But entrants must also fear that if the reservation wage is low, then firms have an incentive to "churn" and hence dismiss the entrants before they turn into insiders. If entrants fear churning and thus cannot be sure that they will stay with the firm, then their reservation wage will be higher to make up for this risk. Under these circumstances, firms may find it profitable to offer an entrant wage that is high enough to make clear that the firm will not have an incentive to churn. This entrant wage will need to be higher than a reservation wage under bonding, but be lower than the reservation wage under churning.
Finally, insiders may use their market power to influence not only their own wages, but those of the entrants as well. In the union literature it is often assumed that insiders bargain over both sets of wages, but this feature is not limited to unionized firms. At first sight, it may appear that when the insiders’ positions are protected by labor turnover costs (and possibly also by seniority rules), both the insiders and their employers have an incentive to set the entrant wages as low as possible (equal to the entrants' reservation wages), for then the resulting economic rent (to be shared between the insiders and their employers) is maximized (Frank and Malcomson, 1994; Gollier, 1991). But insiders know that low entrant wages will encourage greater employment of entrants. If entrants turn into insiders in the future, the future insider workforce will expand. Thus, if insiders and entrants are substitutes in the production process, the current insiders could become worse off in the future: either their future wages will fall, or firms will be tempted to dismiss their high-wage senior employees and retain their low-wage junior employees (Solow, 1985). To avoid such adverse contingencies, the current insiders have an interest to push entrant wages above the entrants' reservation wage, thereby restricting the firms' employment of entrants.

For these various reasons, while labor turnover costs raise the insider wage, the entrant wage – unlike the reservation wage – does not fall by an equivalent amount in present value terms. Thus, entrant wages are also raised above the reservation wages. Taken together, the labor turnover costs raise the present value of firms' wage costs of the marginal worker, and lead firms to reduce employment. Outsiders would prefer to find employment at the entrant wage rather
than to remain unemployed, but since firms keep their employment below what they would choose if paying reservation wages to entrants, outsiders are unable to find jobs. The outsiders' choice sets are thus inferior to those of the insiders both at any point in time and over their lifetimes.

**How Labor Turnover Costs Affect Employment and Unemployment**

Labor turnover costs discourage firms from hiring when labor demand rises and from firing when labor demand falls. Employment inertia increases, so that firms' current employment depends more strongly on past employment. In the aftermath of booms, employment will tend to remain relatively high. However, in the aftermath of recessions, employment will tend to remain lower than it would otherwise have been.

Other implications concern working hours and capital-labor substitution. Since labor turnover costs are usually associated with replacement of employees rather than variation in their hours of work, labor turnover costs tend to encourage variations in working hours rather than in people employed. Furthermore, the greater are these labor turnover costs, the greater will be firms' incentives to respond to cyclical fluctuations through capital-labor substitution rather than through hiring and firing. Consequently, the more pro-cyclical will the swings in productivity be.

These various effects operate over the short- or medium term. Over the long run, the influence of labor turnover costs on the level of employment is less clear-cut. Since labor turnover
costs discourage both hiring and firing, the direct effect of the labor turnover cost on long-run employment could be positive or negative. The literature in this area indicates that this effect on employment depends on such factors as the discount rate, the stochastic process generating the demand shocks, the rate of productivity growth, and the quit rate.

For example, when demand shocks are permanent, then firing costs may stimulate average employment, because while a firm that fires incurs the firing costs immediately, a firm that hires faces the firing costs only in the uncertain future. Thus firing may be discouraged more than hiring (Bentolila and Bertola (1990)). Conversely, hiring and training costs tend to reduce average employment, because when a firm hires, it incurs the hiring costs immediately, but when it fires workers such costs are uncertain and occur in the future. If the firing costs are more important than the hiring and training costs – an assumption often made implicitly, without obvious justification, in this branch of the literature – these particular mechanisms may raise rather than reduce employment over the business cycle.

However, this result can be overturned by other factors. For example, when firms also face idiosyncratic productivity shocks, then a rise in firing costs reduces the number of firms engaged in firing; consequently aggregate firing is discouraged less than would otherwise be the case (Bentolila and Saint-Paul, 1994). Or when the marginal revenue product curve drops quickly, so that the marginal revenue product in a boom is sufficiently low relative to its value in a slump, then a rise in firing costs may also reduce employment (Bertola, 1992). Chen, Snower, and Zoega (1999) show
how that the expected employment effect of firing costs depends on the combination of productivity growth and the probability of recession or boom.\textsuperscript{4}

But perhaps much more important, there is also an indirect effect of labor turnover costs on long-run employment, which occurs through wage formation. As noted in the previous section, an increase in labor turnover costs is likely to raise the present value of firms' wage costs thereby leading to a fall in employment. This effect is unambiguous in its direction and could well be large.

Taking the direct and indirect effects together, the influence of labor turnover costs on employment and unemployment can be shown to depend on the nature of macroeconomic fluctuations. When business cycles are short and shallow (as in the 1950s and 1960s), labor turnover costs encourage labor hoarding in recessions in expectation of an imminent upturn, and thus these costs may have only a small (and possibly positive) effect on employment. But when the cycles are long and deep (as in the aftermath of the two oil price shocks and the recession of the early 1990s), the labor turnover costs may not do much to discourage layoffs in recessions, but they do discourage hiring in boom because of uncertainty about the strength of the upswing and out of fear of another major downturn around the corner. Under these circumstances, labor turnover costs may have an adverse influence on employment (Lindbeck and Snower, 1988b, ch. 11). Diaz and Snower (1999) argue that when macroeconomic fluctuations are highly persistent, the wage-mediated effect of turnover costs on employment is strong relative to the ambiguous direct effect of these costs on employment. Thus, a given set of turnover costs have a more adverse influence on employment
when the fluctuations are more persistent. These medium-run effects may extend over many years, often a decade or more into the future. Many continental European firms, facing relatively high labor turnover costs, have been particularly vulnerable to this phenomenon, illustrated by the high persistence of unemployment in this part of the world.

Another channel whereby labor turnover costs may reduce employment is through the hold-up problem, as surveyed in a labor market context by Malcomson (1997). Labor turnover costs turn hiring into an investment decision, where the employer must face immediate costs in the expectation of a future return (Oi, 1962). Through later wage bargaining, the insiders may be able to capture some of the return to this investment. As result of this hold-up problem, firms may have an incentive to employ fewer people than they otherwise have.

**The Dynamics of Employment and Unemployment**

In addition to examining labor market activity at any point in time during the business cycle, the insider-outsider theory may also be used to analyze patterns of employment and unemployment dynamics. Broadly speaking, the implications cover the behavior of firms, insiders, and outsiders. Let us consider each in turn.

When hiring and firing is costly, firms' current employment level depends positively on their past employment levels. On account of this inertia, the employment effects of temporary labor
market shocks last longer than the shocks themselves, as indicated above. This phenomenon is known as “employment persistence.” As long as the shocks have a stronger effect on employment than on labor supply, employment persistence translates into unemployment persistence. For instance, an oil price shocks leads to higher unemployment even after the shock has disappeared. An extreme form of unemployment persistence is “hysteresis”, when a temporary shock has a permanent effect on unemployment.

Moreover, unemployment persistence may be asymmetric across positive and negative shocks. For instance, since insiders are associated with higher turnover costs than entrants, the degree of employment inertia depends on the insider-entrant composition of the firm's workforce. This composition is determined by the firm's "insider membership rule," describing how employees gain and lose insider status within the firm. Now suppose that the insider membership rule is asymmetric in the sense that insiders who are fired lose their insider status quicker than entrants gain insider status. Then a current macro downturn may have a more persistent effect on future unemployment than does a current macro upturn of equal magnitude (e.g. Lindbeck and Snower, 1987a). The degree of asymmetric persistence depends on the proportion of insiders who move to unemployment rather than to other jobs (the smaller is the proportion of job-to-job movements, the quicker insiders lose insider status) and the importance of firm-specific skills relative to general skills (the more important are the firm-specific skills, the quicker insider status is lost).

Turning to the dynamic behavior of insiders, one form of asymmetric persistence arises on
account of asymmetric employment probabilities. In a boom, when insiders’ jobs are reasonably secure, a downswing may be expected to reduce job security by more than an upswing raises it. Then senior insiders – exerting stronger influence on insider wages than their junior counterparts – may insist on unchanged wages in a downturn (leading to layoffs of their junior colleagues), whereas they may press for wage increases in an upturn. The more stringent the seniority ranking, the more asymmetric will be the resulting employment persistence.

Moreover, insiders’ incentives to prevent wage underbidding by threats of non-cooperation and harassment may be asymmetric over the cycle. Insiders are more likely to be exposed to wage underbidding during business downswings than during upswings, both from outsiders and from junior workers who in downswings may want to accept wage cuts to keep their jobs.

Whereas these are likely mechanisms, much of the literature on unemployment persistence has gone a different route. For instance, Blanchard and Summers (1986) and Gottfries and Horn (1987) argue that a negative, mean-reverting productivity shock leads firms to fire some of their insiders, and thereby raises the expected job security of the remaining insiders (since the shock is expected to reverse itself). In response, insiders raise their wages and consequently discourage future employment, thereby generating employment persistence. This argument is frequently combined with another, namely, that in the presence of diminishing returns to labor, the insider wage may depend inversely on the size of the insider workforce. Thus, a current contraction of the insider workforce is associated with a higher marginal product of labor than heretofore, allowing insiders to
achieve a higher negotiated wage and discourage future employment. Blanchard and Summers argue that such mechanisms lead to hysteresis.

In our opinion, however, the pervasive emphasis on these mechanisms – particularly in empirical tests of the insider-outsider theory – is misplaced. A combination of hysteresis and random labor market shocks leads to the counterfactual prediction that unemployment hits zero or 100 percent in finite time. Moreover, in practice, productivity shocks are generally not mean-reverting, so the fact that some insiders are fired does not necessarily increase the security of the remaining insiders. In addition, diminishing returns to labor are at best a short-run phenomenon and, in the presence of excess capital capacity, may well be irrelevant to wage determination because when there is excess capital capacity, firms generally vary labor and capital services simultaneously in response to shocks, rather than varying labor relative to a fixed stock of capital in use (Lindbeck and Snower, 1994). Finally, even in the presence of diminishing returns to labor, wages need not be inversely related to the size of the insider workforce. The reason is that when wages are the outcome of bargaining, they commonly depend both on workers’ productivity and on their reservation wages (that would make them indifferent between work and leisure). Whereas productivity falls with employment under diminishing returns, the reservation wage rises with employment if the labor supply curve is upward-sloping; thus the relation between the negotiated wages and employment is ambiguous.

Regarding the dynamic behavior of outsiders, the deterioration of their skills may generate
unemployment persistence. Since firms often exercise some bargaining power in wage negotiations and thus capture some of the economic rent generated by workers' skills, these firms get more rent from skilled than from unskilled workers. Thus, if unemployed workers see their level of skills deteriorate, firms will find them less desirable to hire. In this way, current unemployment gives rise to future unemployment. Unemployment persistence can also arise when long-term unemployed workers are stigmatized by firms (for example, Blanchard and Diamond, 1994). Such stigmatization arises when firms use the length of workers' unemployment spells as a predictor of their potential productivity. Once again, current unemployment comes to depend positively on past unemployment. Finally, unemployment persistence may arise when outsiders' efforts at job search decline with their duration of unemployment. This may occur because their subjective probabilities of finding jobs become smaller, the longer they are unemployed (Layard and Bean, 1989).

Since the outsiders' deterioration of human capital, stigmatization, and depressed job search become more pronounced as the duration of unemployment lengthens, these considerations help explain why real wages and inflation tend to be more responsive to changes in short-term unemployment than to changes in long-term unemployment. Moreover, outsiders' employment probabilities are likely to respond asymmetrically to business fluctuations; that is, the possibility of employment rises in upturns, but doesn't get any lower than zero in downturns. As a result, the resulting employment persistence may be asymmetric.

Many of the lagged adjustment processes discussed above – along with a host of others⁶ - are
complementary. For instance, when there is a temporary, adverse labor demand shock, costs of hiring, training and firing imply that employment will remain low for some time after the shock has disappeared. But because employment remains low, some of the people who lost their jobs become long-term unemployed. If these people are less effective at competing for jobs than the short-term unemployed, wages will be higher than they would otherwise have been. Consequently, employment remains low for far longer. When adjustment processes are complementary along such lines, the joint influence of the network of adjustment processes is greater than the sum of the individual processes, and thus it may take unemployment a long time to approach its long-run equilibrium in the aftermath of a shock. Karanassou and Snower (1998, 1999) offer a theoretical and empirical assessment of such complementarities.

Thus, in analyzing the movement of unemployment, the insider-outsider theory shifts the focus away from exclusive concentration on changes in the long-run unemployment equilibrium, and emphasizes a complex system of interacting adjustment processes and related unemployment persistence. The adjustment dynamics of the insider-outsider theory have two important implications, particularly in many European countries.

First, the longer-term swings in unemployment, may be attributable not just to movements in the long-run unemployment rate, but also to prolonged deviations of actual unemployment from the long-run rate (Lindbeck, 2000). For example, Henry, Karanassou and Snower (1999) suggest that the movements of U.K. unemployment over the past 30 years are due
largely to slow adjustment dynamics. By the same token, the fall of the U.S. unemployment rate over the 1990s may be due not just to a fall in the long-run equilibrium unemployment rate (e.g. the NAIRU), but also – perhaps largely – to the presence of favorable shocks (falling oil and other raw material prices and increased international competition) combined with persistence mechanisms.

Second, when the labor market is growing, labor market adjustment processes may prevent the unemployment rate from converging to what the long run unemployment rate would be in the absence of adjustment costs. When the labor demand and labor supply curves are continually shifting outwards, the adjustment processes may never have a chance to work themselves out fully. Under these circumstances, the adjustment processes determine how far employment and the labor force are lagging behind their moving targets, even in the long run (Karanassou and Snower (1998)).

**Labor Market Segmentation**

The insider-outsider theory has important implications also for the nature of labor market segmentation. It is usually assumed that the “primary” and “secondary” sectors differ in terms of (a) wages: whereas secondary-sector employees receive their reservation wages (at which they are indifferent between employment and unemployment), primary-sector employees earn significantly more; (b) job tenure: primary-sector employees have significantly higher
probabilities of retaining their jobs than do their secondary-sector counterparts; and (c) promotion opportunities: the primary sector has promotion ladders whereas the secondary sector operates more like an auction market. (In practice, of course, the distinction between the primary and secondary sectors is one of degree rather than kind.) Whereas the efficiency wage theory rationalizes this distinction in terms of asymmetric information – primary-sector employers use their wage offers to stimulate their employees’ productivity, while secondary-sector employers do not – the insider-outsider theory rationalizes the distinction in terms of labor turnover costs. Specifically, these costs are significant in the primary sector but not the secondary sector, and thus primary-sector employees have some market power whereas the secondary sector is competitive.

To the extent that labor turnover costs rise with job tenure within the primary sector, the insider-outsider theory can also help explain the presence of tenure-related wage scales. We are not arguing, of course, that labor turnover costs are only factor here. Human capital theory predicts that the slope of intertemporal wage scales depends on how workers' productivities change with their length of job tenure (Becker, 1962). In addition, efficiency wage theory predicts that firms set a pattern of rising wages with greater job tenure to motivate their employees – a kind of bonding contract (for example, Lazear, 1981). However, since labor turnover costs (as a function of job tenure) are generally not perfectly correlated with productivity and incentive effects (again as a function of tenure), the insider-outsider explanation is, at least in principle, different from
explanations in terms of the human capital and efficiency wage theories.

Insider-outsider theory is also able to account for the observation that some firms pay their workers higher wages than other firms do – regardless of the occupations, ages, and tenure groups from which the workers are drawn. Numerous studies have found that industries that pay comparatively high wages tend to have relatively high profits, high concentration ratios in product markets, high capital-labor ratios, and high union density (Dickens and Katz, 1986). The insider-outsider theory sheds some light on these patterns by indicating that insiders’ wages are the outcome of a process whereby insiders and their employers share the available economic rents. Lindbeck and Snower (1990a) show that, under a broad range of conditions, firms stand to lose more from a breakdown in wage negotiations the greater the profit opportunities available under agreement, the greater the capital-labor ratio of the firm, and the greater the concentration ratio of the industry. Thus, employers in such settings are more willing to pay high wages to insiders. Moreover, the greater is union density in an industry, the more leverage unions are able to give insiders in their threats of obstructive activity under bargaining disagreement, and thus the higher the resulting insider wages – the well known “union premium”.

**Institutions**

The insider-outsider theory has significant implications for various labor market institutions. In this section we consider two: labor unions and social norms. (Implications for policy institutions
are considered in the final section.)

**Unions**

In much of the traditional literature of union behavior, the existence of unions is taken for granted, rather than explained. The union is usually assumed to maximize the sum of its members' utilities or the expected utility of a representative member. This means that the union is viewed as maximizing a welfare function that depends positively on the wage and employment level. This analysis needs to be modified when the insider-outsider analysis is applied to unionized versus non-unionized workers.

First, the insider-outsider theory provides an explanation of what gives unions their clout: Firms are reluctant to replace their high-wage unionized employees by low-wage non-unionized ones because of labor turnover costs, including costs related to the powers of insiders to intimidate underbidding outsiders, for instance, when the latter try to cross a picket lines. The claim of the insider-outsider theory is that unions’ clout derives largely from this source, for if firms found it costless to replace all their high-wage unionized employees by low-wage non-unionized ones, they would do so.

Second, by implication, the insider-outsider theory highlights the idea that indifference curves of a union in wage/employment space have a kink: they are downward-sloping as long as employment is less than union membership, and flat when employment exceeds membership
Thus in a cyclical upturn, when employment often exceeds membership, the union will push for wage increases; but in a downturn when employment often falls short of membership, the union will accept a combination of wage restraint and employment cuts.\footnote{Carruth and Oswald, 1987}

Third, insofar as labor turnover costs are the source of unions' market power and these costs are generally positive but not prohibitive, the insider-outsider theory raises doubts about the monopoly union model, because a union can be a monopolist only if labor turnover costs are prohibitive. It also raises some doubt about the standard union bargaining models in which unions and employers have bilateral monopoly power. Such bilateral monopoly power can arise only if labor turnover costs are so high that the unionized insiders are more profitable than the non-unionized outsiders at any feasible insider wage. In practice, turnover costs are generally not that high, and thus employers and unions must take competition from outsiders into account in their wage bargaining, depending on the size of labor turnover costs.

Under these circumstances, union wage setting depends on more than union’s objectives, firms’ employment decisions, and relative bargaining strengths. It is also influenced by additional constraints, such as a “relative profitability constraint” whereby the union wage is no greater than the outsiders’ reservation wage plus the relevant labor turnover costs, and a “credible threat constraint” whereby the union wage must be such that if firms reject the unions’ wage proposals, union members have an incentive to do what the union has threatened (e.g. go on strike). (For details, see Lindbeck and Snower (1987b).) Though the underlying claim – that labor turnover costs
are central to the exercise of union power – has far-reaching implications for modeling union behavior, the union literature to date has devoted little attention to the subject, except for the recent notion that the labor-demand curve may be kinked.

Fourth, the insider-outsider theory helps clarify the conditions under which unions arise. Basically, in the insider-outsider framework the existence of unions is explained by their ability to further the interests of the insiders, because unions are able to augment labor turnover costs. In particular, firms are more likely to grant wage increases when the alternative is the replacement of all their unionized employees than when the alternative is just firing a single employee. Unions also provide new tools of rent seeking, such as strikes, work-to-rule activities and picket-lines, which tend to be effective only when workers act in unison. Finally, unions also act as interest groups in the political process, lobbying for job security legislation and other sources of labor turnover costs. Unions, of course, also fulfill other important functions, such as protecting individual workers against arbitrary and discriminating treatment by employers, transmitting information about the production process between employees and management and participating in the political dialogue in society at large, hence contributing to political pluralism.

According to the insider-outsider theory, unions are likely to thrive when (a) individual employees' turnover costs are high, so that there is a significant payoff from coordinating their rent-seeking activities; (b) firms have significant market power in product markets, so that there are rents to be exploited; (c) the political process is susceptible to job-protection lobbying; and (d)
employment legislation protects rights to strike, picket, and other union rent-creating activities. This means that unions may be regarded both as a consequence and a cause of labor turnover costs.

Fifth, the insider-outsider theory challenges the traditional assumption that unions maximize the sum of the welfare of all union members or of a representative union member. Since labor turnover costs rise with job tenure, the interests of senior union members are substantially different from those of the junior ones. In particular, senior insiders tend to have higher probabilities of retaining their jobs than do their junior counterparts, and thus the former group – often the more influential within the union – will tend to push for higher wages than the latter. Similarly, when unions have both employed and unemployed members (a common occurrence in Europe, for example), the interests of the employed members diverge from those of the unemployed ones. In the presence of labor turnover costs, the probability that an unemployed person is hired is significantly less than the probability that an employed person will be retained.

In practice, unions are primarily concerned with their employed workers, partly because the employed are more numerous. Indeed, in many countries it is common for the unemployed eventually to drop out of their unions. This observation suggests that when employment falls short of union membership, the unemployed tend to leave the union, reducing union membership; and when employment exceeds membership, the new entrants tend to join the union. Union membership dynamics – which have recently received considerable attention in the literature – are important because, as noted, they influence the path of wages and employment over the business cycle.
Finally, insofar as labor turnover costs rise with job tenure, the insider-outsider theory suggests why unions typically support seniority arrangements in which wages rise with job tenure. In the context of the insider-outsider theory, the rate at which retention probabilities and wages rise with job tenure depends on the rate at which labor turnover costs rise with tenure.

**Social Norms**

The insider-outsider theory explains why jobs aren’t given to the lowest bidders even though all agents display purposeful (rational) economic behavior, as usually assumed in economics. But there are other, complementary, ways of looking at the issue. For sociologists and many other observers of human behavior, it may be more natural to argue that there is a social norm against outsiders underbidding the wages of incumbent workers, and against the acceptance of such lower bids by firms. Although economists have largely neglected social norms, at least until recently, this concept has for a long time been a basic one in sociology, in particular in the tradition of Parsons (1952). Social norms imply that a certain type of behavior is expected by others, and enforced by external sanctions that are social rather than economic in nature. Conformity with social norms is met by approval, status, and pride, and deviation is met with by disapproval, stigmatization and shame.

There is ample evidence that much human behavior is strongly influenced by social norms (Coleman, 1990). Sociologists generally agree that the labor market is a hotbed for social norms: for
instance, in the case of the determination of "proper" work effort and "fair" relative wages (for
instance, Elster, 1989, in this journal).

Indeed, some scholars have argued that rational economic calculations are not sufficient to
generate and maintain the type of non-cooperation and harassment behavior postulated in one central
version of insider-outsider theory, and that a broader social norm against wage underbidding is in
fact required (Akerlof, 1980, 1991; Elster, 1989). In particular, threats of non-cooperation and
harassment by insiders have been alleged to be time-inconsistent (Fehr, 1990). The argument is that
the insiders’ refusal to cooperate with underbidders may reduce the productivity not only of the
underbidders but also of the insiders themselves, and that harassment may create disutility for the
harassers themselves. Then non-cooperation and harassment threats may not be credible.

These assertions generally do not hold water when insiders threaten to harass colleagues who
do not refuse cooperate with or do not harass underbidding outsiders. Nor do they hold in a multi-
period context, for insiders may very well be willing to pay a price today to discourage underbidding
in the future. Non-cooperation and harassment would then viewed as an investment in credibility.

But even though the non-cooperation and harassment version of the insider-outsider theory
does not require a social norm against wage underbidding, the model becomes richer and more
powerful if combined with the notion of a social norm against wage underbidding. Such a norm may
be particularly effective in constraining behavior when the norm is shared not only among
incumbent workers and potential underbidders, but among managers as well. Some questionnaire
Assar Lindbeck and Dennis J. Snower

studies suggest that managers have even stronger aversion than outsiders to pursue, or accept, underbidding offers (Agell and Lundborg, 1995; Bewley, 1998). Insiders’ labor market position may also be boosted by a social norm among employers according to which lay-offs and firing should be avoided as much as possible even in business downturns – a norm that has perhaps weakened in the last decade (Osterman 1999).

Indeed, rather than thinking of social norms as rendering the non-cooperation and harassment version of the insider-outsider theory unnecessary, this theory may be viewed as providing an explanation of how a norm against wage underbidding could emerge and be sustained. Since it is in the insiders' personal self-interest to protect themselves against underbidding, they have incentives to serve both as "senders" and as "monitors" of such a norm. Unions may be important in this respect, not only by making non-cooperation and harassment activities more efficient but also by helping establish and monitor social norms against underbidding among workers and managers.

**Empirical Evidence**

The insider-outsider distinction is frequently invoked today in discussions of the employment experiences of developed countries. For instance, higher persistence of unemployment in Europe than in the United States is often seen in the perspective of the insider-outsider theory, although often the terminology and conceptual framework of the theory is only informally applied. Above we have noted other practical applications of the theory, e.g. the unemployment effect of
labor turnover costs over different phases of the business cycle and the role of unemployment persistence versus changes in the long-run unemployment equilibrium rate in explaining unemployment movements. Despite interpretative problems, such informal applications are an important route, in conjunction with other evidence, for judging its realism and usefulness. Moreover, strategic assumptions of the theory have close real-world correspondences. Hiring, training and firing costs – as well as host of other labor turnover costs – are pervasive, and historical experience documents insiders’ hostile treatment of outsiders who try to underbid insider wages; terms like “scabs” reflect insiders’ attitudes to such outsiders.

Systematic confrontation of the insider-outsider theory with empirical evidence by way of formal statistical tests have also been carried out. One set of empirical tests has considered the extent to which real wages are affected by conditions inside individual firms, and not only conditions outside. "Inside factors" usually considered are the productivity of workers, output prices or profits, hiring and firing costs, and the bargaining strength of workers. (The idea that profits influence wages is an old one; see, for instance, Slichter, 1950). Obvious "outside" influences are factors affecting the outside options of workers, such as the levels of unemployment, unemployment benefits and other welfare-state benefits, as well as wages offered by other firms. Contrary to what is implied by some authors (e.g. Layard and Bean, 1989), the finding that outside influences are relevant to wage formation is of course not evidence against the insider-outsider theory. The theory predicts that not only outside, but inside
and outside, influences are significant, since outsiders affect the wage setting process indirectly via their influence on insiders’ fall-back positions and outside options in wage bargaining.

Studies along these lines have been done in many contexts: cross-section regressions over countries and production sectors (Coe, 1990; Holmlund and Zetterberg, 1991); across regions such as U.S. states (Kendix, 1991); and using micro-data for individual firms (Blanchflower, Oswald and Garrett, 1989; Nickel and Wadhwani, 1990). Holmlund (1990) and Lever (1995) offer summaries of parts of the empirical literature in this field, and Gregory (1986) explores the empirical implications of inside factors for the Phillips curve.) Most tests are consistent with the hypothesis that both inside and outside factors influence real wages. When wages are largely set in the interest of insiders, as postulated by the insider-outsider theory, we would also expect that lay-off rates have a negative influence on real wages, since higher lay-off rates threaten the insiders’ jobs. A study for the Netherlands by Graafland (1992) is consistent with this prediction.

The implications of different degrees of “insiderness” for wage formation have also been studied empirically. For instance, Dolado and Bentolila (1993) have found evidence that an increase in the number of fixed-term employees, who are workers with weak inside status, boosts the market powers and the real wages of permanent workers who have the highest inside status. Regarding the influence of different degrees of "outsiderness" on wage formation, many studies indicate that the long-term unemployed exert considerably less (downward) pressure on real wages than do the short-term unemployed (e.g. Layard and Nickel, 1987). While this observation is consistent with the
predictions of models lying in areas of overlap between the insider-outsider, human capital, and search theories, the prediction that (un)employment inertia increases with the level of labor turnover costs is more specific for the insider-outsider theory, and it is broadly consistent with empirical studies (Holmlund, 1990; Nickell and Layard, 1997, Table 15). This observation also fits well with empirical evidence that countries with high labor turnover costs have lower exit rates from the unemployment pool than other countries (Alogoskoufis et al., 1995).

Another important empirical issue concerns the influence of labor turnover costs on the average rate of unemployment over the cycle. While theoretical considerations above indicate that this influence is ambiguous, the insider-outsider theory adds that the greater is the market power of insiders, the more likely is the influence to be negative.10

Moreover, the insider-outsider theory leads us to expect that the frequency and duration of unemployment spells for typical outsider groups such as young workers, women and some minorities will be comparatively high in countries where insiders enjoy relatively high job security and strong market power. The reason, clearly, is that in the presence of high labor turnover costs, insiders can insulate themselves from macroeconomic fluctuations to a considerable extent, leaving outsider groups to bear the main burden of negative shocks. We also expect that across countries with about equally high legislated labor turnover costs but different cyclical fluctuations, the unemployment rates of typical outsider groups will differ widely whereas the unemployment rates of the core groups will be more uniform, and relatively low. The available empirical evidence is
consonant with these predictions (Elmeskov, Martin and Scarpetta, 1998, Fig. 3).

Some of the theory's predictions about labor-market dynamics have also been tested. For instance, the observation that unemployment persistence is higher in most countries in Western Europe than in the United States is consistent with the insider-outsider theory, which predicts that persistence would increase with higher labor turnover costs.

Other studies of labor-market dynamics refer to the relations between short-term (cyclical) macroeconomic fluctuations and labor hoarding, hours of work, capital-labor substitution and productivity movements. For instance, in countries with high labor turnover costs, we would expect relatively large labor hoarding during recessions. This pattern seems to be the case; the size of the workforce fluctuates less relative to output in Western Europe than in the United States over the business cycle. This finding helps explain why labor productivity is more procyclical in Western Europe than in the United States. Since labor turnover costs are related to variations in the number of workers, we would also expect that hours of work fluctuate more in countries with high labor turnover costs. This prediction is consistent with data for West Germany and the United States (Abraham and Hauseman, 1993).

Some evidence also exists that wage adjustment is asymmetric in the sense that real wages rise more readily in upturns than they fall in downturns, as predicted by various dynamic insider-outsider models. For example, Holzer and Montgomery (1990) find that wages adjust asymmetrically with respect to sales growth, according to a survey of U.S. firms in 1980 and 1982.
Blanchflower (1991), studying the British Social Attitudes survey, finds that wages rise when workers expect employment at their firms to rise, but wages remain constant when they expect employment to fall. Nickell and Wadhwani (1990) also find evidence of downward wage rigidity for firms surveyed in the United Kingdom. Begg et. al. (1989) find evidence of asymmetric persistence in the United Kingdom, Japan, and (to a lesser extent) Germany.

Many empirical studies have also tested the predictions of the insider-membership variant of the insider-outsider theory (e.g. Blanchard and Summers (1986)), according to which wages depend inversely on past employment, and the current unemployment rate is so persistent that it is the best predictor of the future unemployment rate (the hysteresis phenomenon). We have argued above that this variant, though predominant, is an unlikely special case and indeed there is little empirical support for it (Holmlund, 1990; Lever, 1995; Nickel and Wadhwani, 1990).

High labor turnover costs are expected to boost labor productivity in the long-run perspective, though perhaps at the expense of the employment level, by stimulating higher real wages and thus encouraging capital-labor substitution. The insider-outsider approach suggests that this effect may occur even with plenty of unemployed workers, because high labor turnover costs protect the insiders' position. Investment in firm-specific human capital is also expected to be encouraged by high labor turnover costs. These may be two of the reasons (among many) why western Europe experienced a higher rate of long-run productivity growth than the United States from the 1950s through the 1980s. But this boost in measured labor productivity happens because
low-productivity workers are being kept out of work by high real wages due to the power of insiders. Of course productivity growth depends on many factors in addition to those above, as evidenced by the higher productivity growth in the U.S than in Western Europe during the second half of the 1990s.

Regarding some political-economy implications of the insider-outsider dichotomy, there is some support for the claim that insiders, who constitute a majority of the electorate, are able to push through (or at least support) legislation boosting the costs of firing workers, facilitating strikes and extending collective bargaining agreements to firms without organized workers (for instance, Saint Paul, 1996). There seems also to be a positive correlation between the strictness of employment protection legislation for permanent workers and the so-called "excess coverage" of wage contracts, expressing the extent to which union wage agreements are extended to non-union members. One interpretation is that insiders, who benefit from strict employment protection legislation, have been able to insist on legally enforced extension of wage agreements as a protection against wage underbidding (Elmeskov, Martin and Scarpetta, 1998).

Taken as a whole, these assorted pieces of evidence – informal applications as well as formal tests – lend support to the proposition that the insider-outsider approach helps account for a number of salient empirical labor market phenomena. There is, of course, room for a great deal more empirical work in this area. In particular, some of the evidence, while illustrating that the insider-outsider approach fits with recognized economic patterns, is also consistent with other labor market
theories. Given the centrality of labor turnover costs to the insider-outsider theory, it is first of all important to measure labor turnover costs with better precision. It would then be easier to test a number of implications of the insider-outsider theory. We would like to know whether the magnitude of insider wages (relative to the reservation wage), and the share of wages relative to profits, tend to rise with labor turnover costs. This would also help us judge the relative importance of "inside" versus "outside" factors in wage formation, and the extent to which union power and union wage premiums depend on labor turnover costs. It would also be important to explore if the relative degree of employment persistence, across firms or sectors, depend on the relative magnitude of labor turnover costs. It would also be worthwhile to examine the importance of labor turnover costs for the intersectoral wage structure, as well as for labor market features like job queues, retention rates, and the consequences of labor turnover costs for employment variability. Moreover, a lot more work needs to be done on the empirical evaluation of asymmetric persistence, which may shed light on the question why, over the past 25 years, European unemployment has tended to ratchet upwards from one recession to the next while U.S. unemployment has remained essentially trendless.

As this list of topics makes clear, and it is easy to extend, empirical investigation of insider-outsider phenomena is still in its infancy.
Policy Implications

Although the policy implications of the insider-outsider theory are diverse (Lindbeck and Snower, 1988b, ch. 11; 1990c), they have a common thrust: Insofar as insiders have more favorable opportunities than outsiders, policies that create a more level playing field in the labor market can improve both efficiency and equity. Either "power-reducing policies" that mitigate the insiders' market power relative to that of outsiders, or "enfranchising policies" that make outsiders more attractive to employers and give them a stronger voice in the wage determination process, can help level the playing field (Lindbeck and Snower, 1990b).

Power-reducing policies include restrictions on strikes and picketing and relaxing job security and seniority legislation (for example, through laws to liberalize firing procedures, reduce litigation costs, and reduce severance pay) – at least in countries where such legislation is particularly strict. However, even when these policies improve equity and efficiency, they are usually not Pareto-improving, since they tend to reduce the welfare of insiders. Thus the insiders may resist these policies either through the political process or through workplace actions.

Enfranchising policies can take various forms: vocational training programs and job counseling for the unemployed; employment subsidies (or lower payroll taxes) or tax credits for low-productivity workers; employment vouchers for the long-term unemployed (Snower, 1994); profit-sharing schemes, whereby employees receive part of their remuneration as a share of profits (Weitzman, 1984); schemes to convert wage claims into equity shares (Sinn, 1998); tax and
regulatory policies to reduce barriers to the entry of new firms; and policies to reduce the occupational, industrial, and geographic coverage of union wage agreements. Without such enfranchising policies, power-reducing policies may simply turn some unemployed outsiders into working-poor.

Since the insider-outsider theory emphasizes unemployment persistence rather than just changes in the equilibrium unemployment rate, the theory also points to the risks of pursuing policies that keep aggregate demand substantially below full capacity utilization of labor for long periods of time. In this sense, the insider-outsider theory has important implications for monetary and fiscal policy. A final policy implication of the insider-outsider theory concerns the magnitude of required policy change. “Timid” labor market reforms, whereby policy parameters are changed by only small amounts, are likely to be ineffective overcoming the hurdles of substantial labor turnover costs. The main reason is that substantial labor turnover costs give rise to substantial kinks in firms’ labor demand curves, over which firms have no incentive to hire or fire. Furthermore, as we have seen, many of the lagged labor market adjustment processes generated by turnover costs are complementary to one another. Under these circumstances, there is a case for combining expansionary demand management with "bold" structural reforms to stimulate employment after periods of persistently high unemployment.
The authors are deeply indebted to William Baumol, Brad deLong, Alan Kruger, James Malcomson, Andrew Oswald, Ron Smith, Timothy Taylor and Michael Waldman for profound comments.
References


Holmlund, Bertil. 1990. "Unemployment Persistence and Insider-Outsider Forces in
Wage Determination." Stockholm: FIEF.


Sinn, Hans-Werner. 1998. “Wage Differentiation and Share Ownership to Counter


1 Some aspects of this literature have been surveyed by others like Ball (1990) and Sanfey (1995).

2 In Shaked and Sutton, 1984, firms are constrained to bargain with their insiders for a specified span of time; thus insiders and outsiders are not substitutable during this span, but perfectly substitutable beyond it.

3 We are grateful to James Malcomson for this point.

4 Vetter and Andersen (1994) derive the result that the exercise of insider power raises employment relative to the competitive outcome. To reach this conclusion, they not only assume that entrants receive their reservation wages, but also that (a) all labor turnover costs are production-related (a training cost), (b) they are not manipulable by the insiders, and (c) insiders can reduce their wages to avoid entrants' underbidding, while under competitive conditions these same workers don't do this. Thus their unrealistic result is based on the unrealistic underlying assumptions.

5 An insider membership rule is formally analogous to a capital accumulation equation in which the current capital stock is equal to the capital stock surviving from the previous period plus the amount of investment. The analogy is worth taking seriously, for labor turnover costs turn labor into a quasi-fixed factor of production, like capital (Oi, 1962).

6 Other lagged adjustment processes, for example, are generated by the costs of moving from
temporary to permanent employment and from unskilled to skilled jobs. In addition, of course, there are menu costs and wage-price staggering in the real world.

7 An early analysis of the employment consequence of losing union membership status is Horn (1982).

8. Attempts to integrate social norms in economic analysis of the labor market include Akerlof (1980) and Lindbeck (1995).

9 It is worth emphasizing that this is just one of many possible explanations. The notion of "fair wages," like that of "just prices," is an old one. These notions, along with the dislike of competitors who underbid, often have a range of sources. We are grateful to Will Baumol for noting this point.

10 Nickel and Layard (1997) do not find any systematic relation of this type in aggregate data in their cross-country study, though they report that long-term unemployment is raised while short-term unemployment is reduced. By contrast, Elmeskov, Martin and Scarpetta (1999) do find that the average rate of unemployment is raised by such arrangements.