2011

Reconsidering Institutional Labor Economics

Paul Osterman
M.I.T., osterman@mit.edu
Reconsidering Institutional Labor Economics

Abstract
The author illustrates the utility of institutional labor economics and makes a case for a reconsideration of it. Two recent developments motivate this effort: the rise of New Personnel Economics (NPE) as a significant subfield of labor economics and the substantial shifts in work organization that have taken place since the 1990s. Understanding how and why firms have reorganized work opens the door for a renewed interest in institutional approaches. The author explains that the rules of institutional labor markets (ILMs) emerge from the competition between organizational interest groups—unions, personnel professionals, and the government—and competing views of firms’ objectives—resulting in the rise of ILMs, the slow diffusion of High Performance Work Systems, strategies used to obtain a high level of commitment from workers, the use of contingent employees, and the spread of new promotion rules in response to equal employment opportunity pressures. As such, the role of power and influence in establishing work rules is of central concern, though more conventional NPE considerations also remain important.
INSTITUTIONAL LABOR ECONOMICS, THE NEW PERSONNEL ECONOMICS, AND INTERNAL LABOR MARKETS: A RECONSIDERATION

PAUL OSTERMAN*

The author illustrates the utility of institutional labor economics and makes a case for a reconsideration of it. Two recent developments motivate this effort: the rise of New Personnel Economics (NPE) as a significant subfield of labor economics and the substantial shifts in work organization that have taken place since the 1990s. Understanding how and why firms have reorganized work opens the door for a renewed interest in institutional approaches. The author explains that the rules of institutional labor markets (ILMs) emerge from the competition between organizational interest groups—unions, personnel professionals, and the government—and competing views of firms’ objectives, resulting in the rise of ILMs, the slow diffusion of High Performance Work Systems, strategies used to obtain a high level of commitment from workers, the use of contingent employees, and the spread of new promotion rules in response to equal employment opportunity pressures. As such, the role of power and influence in establishing work rules is of central concern, though more conventional NPE considerations also remain important.

A chestnut in any course on industrial relations includes a review of old debates between the so-called institutionalists and more mainstream labor economists. Lester (1946) and Machlup (1946) debated the role of equilibrium and marginal analysis in labor economics and Lampman (representing what he termed the “empiricists”) and Rottenberg (representing mainstream theory) argued about realism and the standards against which theory should be judged (Lampman and Rottenberg 1956). The local labor market studies of Reynolds (1951), Myers and Shultz (1951), and others critiqued the standard assumptions of mobility and perfect information and these scholars believed that “where frictions, imperfections, and the ‘human element’ in labor markets bulk larger . . . the competitive . . . model frequently gives answers that are incorrect, misleading, or unduly narrow and simplistic” (Kaufman 2008: 287).

By the mid-1960s these disputes seemed to have been settled with the rise of human capital theory and the elaboration of search and information models that accounted for many of the institutionalists’ observations regarding the operation of labor markets. The institutionalists were not eliminated, but they retreated into schools of industrial

* Paul Osterman is the NTU Professor at the MIT Sloan School of Management.

This paper originated as a presentation to a conference organized by Professor Isabel Calderon Gutierrez at the University Carlos III in Madrid and the author is grateful to her and to the participants at that conference for their support. Helpful comments on later drafts were provided by Lucio Bacarro, Frank Dobbin, Robert Gibbons, Sandy Jacoby, Thomas Kochan, Mike Piore, and participants at the SASE annual meeting in Costa Rica, where an earlier version of the paper was presented. The author also thanks Peter Auer, Jerome Gautie, Harry Katz, David Marsden, and Wolfgang Streeck for their feedback.

Industrial and Labor Relations Review, Vol. 64, No. 4 (July 2011). © by Cornell University.
0019-7939/00/6404 $05.00
relations and, less commonly, into business schools. This outcome notwithstanding, my goal in this paper is to reconsider institutional labor economics (ILE) and make a renewed case for it.

One stronghold of institutional thinking has been to research and explain the personnel practices of firms and, in particular, how they organize the rules of their internal labor markets (ILMs). Understanding ILMs is an important arena for institutional theory because, at least as a first approximation, the very existence of administered employment rules seems to point to a different understanding of labor markets than that which emerges from invisible-hand models of supply and demand. In addition, the current period is witnessing a remarkable number of changes in how firms organize work. Understanding what is driving these transformations is central to the intellectual agenda of the field, and this in turn at least raises the question of what value an institutional perspective can provide.

Related to this point is that in the past several decades, the standard templates for organizing work have lost their hold and we now observe substantial diversity in ILMs. This diversity is not difficult to understand if firms face different constraints; however, there is also diversity among firms that appear to be in the same economic circumstances, and this pattern suggests a role for institutional considerations.

Despite the seeming appeal of institutional factors for understanding these patterns in recent years, in a replay of the old pattern, New Personnel Economics (NPE) has emerged as a significant subfield of labor economics and has a tendency to be imperialistic regarding its claims (Lazear 2000). This line of research seeks to explain the personnel and ILM practices of organizations from an optimization perspective. For example, NPE has addressed the general question of why ILMs exist, and has studied optimal compensation and promotion practices, retirement policies, and many other human resource topics. Because NPE is congruent with most of standard economic theory, it has gone unchallenged. At the same time, the older tradition of institutional labor economics, appropriately updated, takes a different perspective, has a great deal to say on the same topics, and is worth a serious reconsideration. This is particularly true because somewhat below the radar screen more institutionally oriented scholars, many of whom are sociologists but some of whom are more direct descendants of the older institutionalists, have enlarged and deepened their own analysis of organizations. Hence the substance of institutional economics is deeper and more vital than has been heretofore understood.

The Argument

There are three logical steps in a research program aimed at understanding ILMs. The first is to ask whether organizations have rules that at least nominally shape employment practices and outcomes. The second step is to determine whether these rules are binding in the sense that they lead to outcomes that are different than those predicted by standard supply and demand models. Assuming that the answers to the first two steps are positive, the third step is to explain the origins of those rules and why they persist.

1 There is considerable empirical work that demonstrates diversity. At the anecdotal level, examples include the contrast in employment practices between Walmart and Costco (Greenhouse 2005; Cascio 2006), the divergence of Southwest Airlines from the practices of other large carriers, or the HRM practices of the software firm SAS compared to its competitors (Pfeffer 1998). More systematic work within industries, such as automobiles and telecommunications, also points to the same conclusion. For example, in their study of telecommunications and automobiles, Katz and Darbishire (1999: 4) reported that “within both union and non-union sectors the extent of variation in wages, work practices, and other employment conditions has increased.” In a review of the literature on call centers, Batt and Moynihan (2002) identified alternative approaches toward organizing work (which they labeled “mass production,” “professional service,” and “mass customization”). Drawing on a survey of 354 American call centers, they found that these models are all represented and the diversity of practices remained even after the sample was divided into centers serving residential and business customers. Additional evidence of the diversity of practices within customer segments is provided in Batt (2002).
The simple existence of ILM rules is accepted by industrial relations scholars and has been so at least since the publication of Doeringer and Piore’s (1972) book on the subject. The second step in the chain of logic is more difficult because of the challenge of establishing the counterfactual (that is, what the results in a more standard labor market would look like) and testing against it. Firm-level data are required because national datasets lack detail on the specific personnel practices that shape outcomes for the individuals who are surveyed. As a result, the research that addresses this question typically studies individual firms although in some instances the data come in the form of descriptive or ethnographic accounts, and in other cases, of econometric analyses of personnel records. The sophisticated econometrics should not, however, obscure the fact that essentially each study deals with a sample of one or just a few firms. In this literature, NPE researchers have played an important role by providing more clarity about where observed patterns diverge from those predicted by traditional labor economics (see for example Baker, Gibbs, and Holmstrom 1994; Gibbs and Hendricks 2004; Gibbons and Waldman 2006). In general, it is fair to conclude that scholars in economics and industrial relations accept the view that ILM rules do in fact shape outcomes in ways that differ from those predicted by models that ignore the administrative rules of the firm.

The debate arises regarding the third question—the origins and persistence of these ILM rules. New Personnel Economics (NPE), as summarized in Lazear and Shaw (2007: 92), “is aimed at modeling firms’ use of optimal management practices.” The prototypical research strategy in NPE is to identify a personnel practice that is empirically prevalent and to develop a model in which the practice in question emerges as the most efficient solution to the problem posed in the model. In other words, practices are seen as flowing from the firm’s effort to maximize efficiency. As a second step, NPE adopts the standard economics equilibrium assumption that if organizations err and adopt sub-optimal practices they will lose out in the competitive market. As a result, natural selection will enforce the diffusion and dominance of practices that flow from the optimizing models.

In developing these models, NPE draws on four theoretical building blocks (Gibbons and Waldman 1999): human capital theory, optimal job assignment models, incentive contracting, and tournaments. The central ideas of each of these theoretical streams are well known. Human capital theory addresses the acquisition of skill and, in the context of a firm, the determination and consequences of learning trajectories. Assignment models try to determine to what extent organizations can best match skills and job requirements. Incentive contracting concerns solutions to the well-known agent-principal problem whereas tournament models focus on the use of promotions to provide incentives and to sort employees by quality.

I present the alternative institutional approach below. In making my argument, however, I want to make it clear that my claim is not that mainstream labor economics does not pay attention to institutions. Whatever merit this claim might have had in the past when the dominant research stream was human capital theory and earnings regressions, it is certainly no longer true in 2011. A rich “standard” literature does examine the impact of minimum wages, unions, employment security legislation, temporary help agencies, and a variety of other labor market institutions. However, in this literature, institutions such as those just listed are viewed as external constraints on the firm and the model of firm behavior—the one exemplified by NPE—remains one of optimizing a well-defined objective function subject to external constraints, whether they are supply and demand or external institutions.

The Institutional Perspective

The central theoretical idea I develop below centers on an alternative perspective regarding firms’ decisions about how to organize work. In addressing this issue, I join others whose contributions have sought to elaborate on the meaning of institutional labor economics and to create space for this

My argument proceeds in two stages. First, organizations are characterized by groups with competing objectives and perspectives regarding the legitimate purposes of the enterprise. Second, the ILM outcomes we observe are the result of an internal political process that is made necessary by the need to resolve these conflicts. In this context, the optimizing pressures emphasized by NPE are important in the political contest, but they are only one of multiple “players”; the actual outcome we observe represents a particular resolution of the conflict. Over time, depending in part on the external economic and political environment, the relative power of different internal constituencies varies, which in turn leads to differences in the relative importance of NPE versus other considerations. In all periods, the group political process is important.

Groups and ILMs

The importance of organized work groups in union and non-union as well as white- and blue-collar settings is well documented in a large and diverse ethnographic literature. Specifically, this literature establishes that these groups act collectively to pressure management to modify practices and policies. According to one literature review (a review that contains numerous citations), “Workers’ goals can include oppositional elements that focus on resistance to management practices” (Hodson 1977: 426). In another review, worker-management interaction is characterized as “managers and workers [acting] jointly to negotiate the trade-offs among management control, worker consent, and labor effort” (Choi, Leiter, and Tomaskovic-Devey 2008: 423). It is important to understand that by “negotiate” the authors refer to informal interaction, pressure, and accommodation, and not to formal union-management collective bargaining. It follows from the nature of these implicit negotiations and political struggles that the outcomes that we observe cannot be presumed to result from optimal decision-making. It is not the case that the competing perspectives described above can simply be modeled as additional constraints to be considered by an optimizing leader.2

There are, of course, other models and ways of thinking that recognize that conflicts exist within organizations over goals and behavior. That point is not new, but what distinguishes the argument here is the emphasis on group rather than individual interests. The NPE approach to conflict focuses on principal-agent issues, but the core of this approach is the view that the individual is the focal actor and that person is seen as an atomistic and selfish utility maximizer looking after his or her particular interests. Conversely, social psychologists emphasize a more expansive view of individual motivation (Pfeffer 2007; Baron and Kreps, forthcoming). As useful as this socio-psychological viewpoint is, however, once again it is important to see that the discussion and examples used to support the argument center on individual-level behaviors such as the role of intrinsic motivation, the importance of reciprocity and gift-exchange, status inconsistency, the anchoring of expectations, and interpersonal comparisons of utility. The perspective I am developing here emphasizes instead group norms, conflict, and negotiation.3

---

2 The idea that organizational behavior/decision-making is not of the optimizing “machine model” mode has certainly been put forward before in the literature by both economists and sociologists. Cyret and March (1963: 31), in the Behavioral Theory Of The Firm, noted that organizations are coalitions and that “basic to the idea of a coalition is the expectation that the individual participants in the organization may have substantially different preference ordering (individual goals) . . . with the obvious potential for group conflict.” However, in the book, the authors emphasize the different goals and coalitions among functional units as they make business decisions (production, accounting, sales, and so on); they devote only two pages (186–87) to employment issues.

3 Other efforts to introduce “social” considerations into standard models are also similarly limited. For example, Robert Solow (1990) argued that the labor market is a “social institution”; however, he basically emphasized variants of efficiency wage models of wage determination and the macro-economic consequence of the fact that individuals respond to fairness considerations and salary norms. He did not discuss groups, group conflict, or competing visions of the objectives of the firm.
This focus on group conflict over goals and rules is consistent with the work of many scholars who have recognized that firms contain multiple groups with quite different perspectives on the legitimate objectives of the organization. Nearly a century ago John R. Commons wrote that “economic conflicts are not merely conflicts between individuals . . . but are conflicts between classifications or even classes of individuals” (quoted in Parsons 1963: 3) In his description of a French bureaucratic organization, Michel Crozier (1967: 298) wrote of “opposing forms of rationality.”

ILM Rules as the Result of a Political Process

The view that ILM outcomes are the result of a political process that aims to resolve group-level conflict stands in contrast to what March and Simon (1958: 37) called the “machine model” of behavior. This is the view that organizations can been seen as algorithms that take inputs, such as data on prices and production processes, and spit out the optimal response, which they then execute. One concern is that this view assumes an organizational computational capacity, doubts about which have long formed the basis of critiques (Cyert and March 1963; Nelson and Winter 1982). More fundamentally, this machine model represents an inaccurate characterization of organizational decision-making and is overly functional. As Jacoby (1990: 317) noted in his critique of standard theory, “There is a tendency to rationalize employment practices in functional or efficiency terms without regard to their historical-causal (how did the practice arise) . . . complexities.”

In thinking about this political process, it is important to understand that although it may be tautologically true that there is a decision-maker who must sign off on a policy, that person does not in fact have the unilateral power to impose an “optimal” decision but rather must take into account the views of constituencies within the organization as well as the power of those constituencies. Once this point is accepted, then the door is opened for an institutional perspective on the determinants of ILM rules.

An extended illustration of the political process perspective occurs in Gouldner’s (1954) classic Patterns of Industrial Bureaucracy. In a case study, he analyzed the origins and implementation of bureaucratic rules regarding issues such as attendance, safety, bidding for jobs, and effort, exploring how the firm created its internal labor market. He described how each rule came into being and argued that they were not based on a universal functional rationality but were instead the result of competing interests. For example, he illustrated how the worldview and interests of several distinct groups of employees as well as both incumbent managers and newly hired outside managers differed from each other and how these differences varied depending on the topic at hand. The point was not just simply that they had different self-interests, but that they also had different views about the underlying rationale of the organization and about what constituted appropriate and inappropriate behavior. As a result, its ILM reflected a political and social process that resolved these conflicts. According to Gouldner, “Rather than assuming that bureaucracy possesses a Gibraltar-like stability, the perspective directs attention into the tensions and problems evoked by bureaucratization” (p. 11).

---

4 Granovetter (1985: 505) made a similar comment in reference to economic modeling more generally: the main thrust . . . is to deflect the analysis of institutions from sociological, historical, and legal argumentation and show instead that they arise as the efficient solution to economic problems. This mission and pervasive functionalism it implies discourages detailed analysis of social structure that I argue here is the key to understanding how existing institutions arrived at their present state.

5 In another important contribution, Zald and Berger (1978) applied social movement theory to organizations and identified three types of political conflict within organizations—coup d’état, bureaucratic insurgency, and mass movements. The first type represents a struggle for power among the elite and is not relevant for the arguments developed here and the third is too episodic to be relevant. However, bureaucratic insurgencies can indeed represent ongoing struggles over the purposes and procedures of the organization that
Dynamics Over Time

The institutional perspective I offer here challenges the view that ILM rules emerge from an optimization calculation. Within organizations, the urge towards optimization and efficiency competes with other objectives held by powerful constituencies. The form that ILMs take—the nature of the personnel rules—can only be understood as flowing from a complex interaction of these objectives and interests. This competition for policy includes an important role for group-level factors such as norms and customs, social structure, competing interests, search for legitimation, and power.6

That said, the viewpoint that I present leaves ample room for NPE-type considerations as represented by constituencies in the organization that place a high value on standard efficiency considerations. In recent years, as competitive pressures on firms have increased, many older ILM practices have seemed more costly and hence the pressure to change them has grown. In addition, NPE scholars might also plausibly argue that underlying technologies have shifted in ways that affect optimal retention policies (e.g., in the direction of more general skills. See Gould (2002)).

All of the above certainly supports the notion that some of what we observe today as ILMs can be understood via standard models as shifting constraints that lead firms to implement new ideas about organizations—what Cappelli (1999) has characterized as “market-in”—such as new uses of on-site suppliers, increased utilization of contractors and outsourcers, and innovations in pay-for-performance. In the empirical material presented in the next section, I begin with a discussion of the history of the emergence of ILMs, and it will be apparent that the power of optimizing considerations has indeed been stronger in some periods than in others. At all times it has remained true, however, that understanding the particular ways in which the economic pressures are refracted through institutional considerations is essential for a full understanding of the origins and operation of the practices that we observe. Groups within organizations seek to redirect pressures to effect change to satisfy their own interests and the actual ILM rules that emerge reflect this institutional fact of life.

Evidence

The arguments developed above point to an understanding of the evolution of ILMs that combines pressures for economic optimization with the kind of institutional considerations that are often overlooked in the literature. I illustrate this perspective by offering five examples: (a) the historical emergence of ILMs and present-day political struggles over occupational boundaries, (b) the slow diffusion of High Performance Work Systems, (c) the strategies of firms to obtain increased levels of effort and commitment from their workforce, (d) the use of contingent employees, and (e) the spread of new promotion rules in response to equal employment opportunity pressures. In discussing each example, my approach will be to identify the political contenders seeking to shape the nature of the ILM practice at hand. The kind of considerations emphasized by NPE models are important,
but they must compete with the exigencies of various internal constituencies. The final resolution—the ILM practices that we observe—are the result of this political competition.

Job Ladders and ILMs

ILMs emerged gradually among American firms from roughly the beginning of the twentieth century until the beginning of the Depression and then accelerated in their diffusion and took on a somewhat different form with the arrival of the New Deal and the rapid growth of union power (Jacoby 2004; Kaufman 2008). Underlying the NPE story is firms’ efforts to find more efficient ways to manage increasingly large organizations as well as their need, which has varied over the business cycle, to limit their turnover costs. Turnover is costly because it entails a waste of human capital, and in a high-turnover organization, provision and investment in on-the-job training is problematic. In order to reduce turnover, compensation is back-loaded, which leads to a system in which people spend long periods of time with the same employer. This effect is intensified by the nature of human capital acquisition. The early investments in training need to be rewarded over time, from the perspective of both the employees and the firm, again implying long careers.

In some versions of the human capital story, there is also a process by which firms need time to learn about the abilities of their employees, which implies firm-based careers (e.g. Farber and Gibbons 1996). The incentive interpretation is somewhat different in that it is less about skills and more about behavior. According to this line of thinking, ILMs help resolve agent-principal problems by creating long-term employment settings in which opportunism can be minimized because there is always another round in the game (Williamson 1975).

The institutional argument regarding the rise of ILMs takes two tacks. The first points to the role of unions in pressing firms to rationalize their employment practices in general as well as to adopt specific practices such as seniority-based promotion. The second refers to the role of the profession, which acts in its own self-interest by encouraging the diffusion of formal employment practices, which then require a personnel staff to administer.

The evidence shows that both the NPE and the institutional stories were important over the course of the twentieth century. At the turn of the century, American firms—even large ones—did not have well-developed internal labor markets. Instead, they managed their workforce via the so-called “drive system,” in which substantial discretionary (often arbitrary) power was put in the hands of the foreman (Jacoby 2004). Turnover was high and thus problematic during periods of growth and labor shortage; this situation led to efforts to stabilize the workforce via creation of job ladders and personnel systems. In his history of the rise of human resource management (HRM), Kaufman (2008: 295) explained that in the pre-New Deal period “HRM moved up and down in response to the threat exerted by unions and government.” Despite this claim, Kaufman placed more explanatory weight on the problems posed by managing large organizations (ibid.: 300). Previously, Jacoby (2004: 102) had written that “employers now found personnel management an attractive alternative, one that promised to simultaneously relieve labor shortages, improve productivity, and promote labor peace.” Moreover, Jacoby related the ebb and flow of the adoption of ILMs to shifting pressures of supply and demand in the external labor market (ibid.: 129). Other correlates included firm size and the stability of their product demand (ibid.: 143).

With the onset of the Depression, the relative importance of the different considerations shifted. In sectors that were unionized, internal political pressures led firms to establish the formal job ladders and seniority systems that characterize industrial ILMs. Non-union firms then imitated these systems to avoid unionization. In addition, personnel specialists within firms actively promoted the formalization of employment practices, a point of view that was clearly in their self-interest. Indeed, during the period spanning the Depression and World War II, there was
very little correlation between the standard optimizing explanations of ILMs—firms’ size and turnover rates—and the adoption of specific formal personnel practices (Baron, Dobbin, and Jennings 1986).

The core conclusion to draw from this material is that when it comes to understanding how the prototypical ILM of the post-war period emerged, the story is considerably more complex than what NPE models imply, and this complexity is consistent with the institutional perspective developed above. There was certainly considerable room for the play of market forces and efficiency-maximizing considerations, but joined to these were the actions of two important interest groups within firms—unions and personnel professionals—as well as the actions of an external institution, the government. ILMs can only be understood as resulting from the interaction of these “actors” as they used their resources to pressure organizations in the directions that they preferred.

The role of internal political struggles over the shape of ILMs is an ongoing story, not limited to a particular historical episode. A powerful illustration of this point is the extensive literature on the determination of boundaries among occupations, the struggle over which is essentially a group-level conflict over the definition of jobs and steps in the ladder of an ILM. At the core of this struggle, and central to the literature, is a rejection of the idea that there is some natural order regarding which tasks are assigned to which professions or occupations (Abbott 1988). The creation of boundaries and steps in the job ladder is, rather, the result of political conflict, some of which is carried out in the public arena via professional societies and licensing authorities, and some of which takes place within the firm. A classic example concerns disputes in the medical field among doctors, nurses, and medical technicians over who can do what for a patient (Freidson 1970), a contest that is ongoing and is fueled by changes in technology and external shifts in funding. Even among doctors there are internal political tensions, such as those between internists and specialists over who can treat which conditions. Abbot (1988) provided a wide range of examples of struggles of this kind in fields such as information technology, law, medicine, engineering, and construction. The literature on how self-conscious groups engage in political struggles within organizations to define the scope of their tasks and to protect themselves against competing groups is not limited to a few elite occupations requiring highly educated workers. Barley (1986) and Nelson and Barley (1997) provided a wide range of evidence of similar processes in a range of “lower level” technical occupations.

The point, for the purposes of this paper, is the ample evidence of self-conscious groups engaged in a power struggle over the shape of job ladders within an organization. It would be difficult to understand the resulting ILM as simply the outcome of an optimizing process by managers who survey the technology and tasks and decide upon the most efficient way to structure work.

The Slow Diffusion of High Performance Work Systems

High Performance Work Systems (HPWS), which involve the use of teams and various forms of quality programs, have been implemented in a range of firms (Osterman 1994; Osterman 2000; Handel and Gittleman 2004) because there is considerable evidence that they lead to higher levels of quality and productivity (Kochan and Osterman 1994; MacDuffie 1995; Ichniowski, Shaw, and Prennushi 1997; Pfeffer 2007). Despite this evidence, their adoption has been slow; this implies, as Pfeffer has observed, that money has been left on the table. Even when adopted, the final form that the work organization takes is often not in line with the initial expectations of management.

The role of internal political struggles between groups of employees and management is apparent in accounts of the adoption of HPWS or the lack thereof. In a case study of non-union and white-collar Hewlett-Packard employees, Beer and Cannon (2004) examined why an effort to implement pay-for-performance compensation
failed. They found that teams of employees resisted because of fears of unexpected changes in their compensation levels; the employees’ resistance took the form of refusing new job assignments and refusing to admit new members to their teams. Beer and Cannon concluded that the process was one of “implicit negotiation” and that “high commitment [work systems] can only be created if employees . . . feel fairly treated and this in turn is a function of how much voice they have” (pp. 14, 17). Other non-union examples include the observation that first-line supervisors resist High Performance Work Organization (HPWO) because they feel threatened by the role of self-managed teams in taking over some of their functions (Coyle-Shapiro 1999).

Case studies in unionized settings lead to the same conclusion. The poster-child is the automobile industry, in which tensions between local unions and management often delayed or reshaped the implementation of HPWS (Kochan, Lansbury, and MacDuffie 1997). More broadly, the literature reports similar issues. Unions can be oppositional when their adversary function is threatened by the more cooperative nature of HPWO systems (Walton, McKersie, and Cutcher-Gershenfeld 1994) and management resists giving up its power to groups of employees (for a useful review of failed HPWO partnerships in unionized settings, see Kochan et al. 2009, Ch. 2).

The case study research is supported by survey methods, although the surveys themselves cannot by their very nature provide as textured a picture. For example, in their survey of the adoption of HPWO practices across a sample of automobile firms, Pil and MacDuffie (1996) found that variables capturing the presence or absence of internal organizational barriers and resistance perform better than simple measures of economic performance. Moreover, as the Hewlett-Packard case makes clear, this pattern extends well beyond the unionized sector of the job market. Whether and how a firm adopts these practices is due to much more than simply an optimizing decision by senior managers.

**Effort and Commitment**

Obtaining higher levels of employee commitment is important in markets in which quality and customer satisfaction are key to competitive success. As a consequence, another recent development is new strategies by firms aimed at addressing this need. NPE approaches employee effort and commitment from the perspective of efficiency wages and incentive compensation models grounded in principal-agent theory. These perspectives are somewhat different in that efficiency wage/shirking models concern level of effort whereas in principal-agent models the issue is that the effort may flow in the wrong direction, that is, in the self-interest of the agent rather than of the principal. What both perspectives have in common, however, is the view that effort is largely decided at the level of the individual and the key to any solution is to get that individual’s incentives correct.

An empirical illustration of this perspective, and evidence that it has power, is Lazear’s (2000) research on Safelite glass. Prominently cited in the literature, the study discusses the consequence of introducing piecework pay systems for auto-glass installers. The research indicated that the installers responded to these incentives and there was a selection process in which the less able employees quit the organization. The point, then, is that people do indeed respond as predicted to financial incentives; at the same time, it is important to remember that the circumstances of Safelite were special. Output was easily measured and monitored and employees worked individually, not in teams or groups. Nevertheless, one can take the point that incentives are important.

Although it is clear that incentives as conventionally understood play a role in eliciting effort, group processes of the sort emphasized by the institutional perspective are also central. The level of effort delivered by the workforce is at least partially determined at the group level and by considerations other than incentive schemes. Consider Toyota or Southwest Airlines, both of which are probably taught in HR courses even more...
frequently than Safelite. These firms are tremendously successful in their industries and both are widely known for obtaining high levels of effort from their workforce. They are also successful in obtaining organizational commitment, an idea that is basically foreign to NPE thinking (Lincoln and Kalleberg 1990). In contrast to NPE expectations, however, both are quite conventional in their incentive systems (and this is as true of Toyota of America as it is of Toyota of Japan) and in both organizations seniority (frequently seen in the NPE literature as the enemy of meritocracy) is important. Why is it that these organizations are able to obtain exceptional effort and organizational commitment while other firms with essentially the same compensation arrangements are not? The answer lies in factors that institutionalists emphasize: group norms and culture.

In thinking about this, it is important to recognize that norms and culture are group phenomena and not examples of the kind of individual-level social-psychological examples discussed above. It is also important to contrast this perspective with the individualist viewpoint typical in the NPE literature. Consider, for example, the comment made by Roberts (2004: 177) in his widely cited book on the internal economics of organizations: “While motivation problems can be manifested at the group level, we focus our discussion on the problems of motivating individuals because these are logically prior and because most of the arguments are more directly made in this context.” This perspective misses the fundamental point that group behavior and group objectives are not simply the aggregation of the behavior and objectives of individual actors.

There is a long tradition in institutional research of recognizing that group culture and norms are important. Seashore (1954) showed that norms operate at the group level, Dunlop (1957: 130) emphasized the importance of custom in both the short- and long-run process of wage determination, and Roy (1959) described how group processes lead to withholding of effort. In its more modern form, organizational culture is a rapidly growing research field (Barley, Meyer, and Gash 1988; Martin 1992).

At both Southwest Airlines and Toyota, employees have developed strong group commitments to the success of the organization (Womack, Jones, and Roos 1997; Pfeffer 1998; Gittell 2003; Rubinstein and Kochan 2001). It is important to avoid romanticizing these organizational cultures, however, for Southwest has had labor trouble on occasion with its pilots and Toyota has worked very hard to keep out the United Automobile Workers. Nevertheless, it is clear to observers (see the citations above) that at the level of the workgroup each organization has succeeded in obtaining effort via broadly shared norms and not simply because of individual incentives or individual gift-exchange.7

There is also a very large ethnographic literature that generalizes the point of the Southwest and Toyota examples, suggesting that the reality is that the workforce essentially engages in a bargaining process with management over effort, a negotiation that can be informal—not linked to formal collective bargaining or the presence of a union. Culture and norms represent the “opening position” of the workforce, and what ensues is implicit negotiation between employees and management over effort.

We have already seen an example of this in the Gouldner (1954) study cited above, in which employee views about the legitimacy of management-specific practices varied; if the practices (ILM rules) were seen as illegitimate, then effort was withheld (pp. 161–62). This pattern was found to be important in a broad range of published ethnographic workplace studies surveyed by Hodson (1999). The ethnographies represented in his survey were undertaken in both unionized and non-unionized factories as well as in white-collar settings such as banks and restaurants. Hodson systematically compiled and coded the reports and

---

7 The understanding that culture and norms are important in eliciting effort may be foreign to NPE but is increasingly shared in the broader economics literature. There remains a tendency, however, to cling to the notion of a decision-maker within the firm who, with this more sophisticated understanding, optimizes the firm’s organizational structure and culture to obtain effort (Akerlof and Kranton 2005).
then examined the relationship between management adherence to work group norms and the degree of effort provided by the labor force. By norms, he meant group-level expectations about fair treatment as well as expectations regarding the organization of work and of production. By effort, he meant what he termed “resistance behaviors” such as acting dumb and sabotaging, as well as “citizenship behaviors,” such as offering ideas, training peers, and providing extra effort. The coded research demonstrated a strong and significant relationship between management adherence to work group norms and citizenship behaviors. This research represents strong evidence of implicit bargaining between groups of employees and management over a core element of the ILM.

**Strategies For Using Contingent Employees**

The use of temporary workers has surged in recent years and is an important element in the fraying of closed ILMs (Autor 2003; Hausman and Osawa 2003). One explanation consistent with the NPE approach is that firms often use contingent work as a strategy for managing the compensation rigidities that result from the persistence of fixed-wage differentials based on work group ideas of fairness (Abraham and Taylor 1996). Related to this is the view that contingent employment enables firms to avoid the constraints that some courts have imposed via restrictions on the employment-at-will doctrine (Autor 2003). What ties these views together is that contingent work is seen as an effort to avoid ILM rules and to return to a more classical labor market. Again, there is certainly good reason to believe that these considerations are important. In practice, however, they are in a political contest with more institutional considerations.

An institutional perspective would begin by observing that the motivations for contingent work outlined above originate in the strength of institutional constraints on the firms’ employment practices and hence the spread of contingent employment in a sense indirectly acknowledges the power of institutional considerations. At the same time, institutional considerations are more than a backdrop against which contingent work is played out: they shape how it is used, that is, the rules that govern its implementation.

An example occurs in the ethnographic research of Smith and Neuwirth (2008), who studied how a temporary help firm interacted with both the HR staff and the line managers of a client firm. How the temps were utilized was the result of a complicated three-way political process within the organization with each interest group having a distinct perspective. The HR group struggled to maintain its role and status, line management sought flexibility, and the temp agency not only searched for business but also negotiated the relationship between the two internal factions. According to Smith and Neuwirth, “What appear to be purely market-mediated employment relationships are constructed in the negotiations and occasional struggles between HR staff, line management, and personnel from labor market intermediaries” (p. 143).

Similar struggles occur between managers and regular employees. When firms use temps they have to pay attention to the reactions of the regular workforce, and the specifics of the ILM rules around temps are often shaped by the need to maintain status differentials. In her study of a Boston area call center and a manufacturing firm, Lautsch (2002) observed that the balance of relations between the two employee groups was a central concern of managers as they decided how to implement contingent work. She noted that “regular workers favored the maintenance of status differences between themselves and temporary workers” (p. 30). In response, management “assert[ed] that [it] would limit the scope of contingent work and its impact on regular jobs” (p. 32). Lautsch added that “management learned . . . that regular workers would rebel if teams were set up so that temporary workers had majority membership, or so that temporary workers were in supervisory roles over regular staff. Gradually these practices were eliminated” (p. 34). In short, efficiency considerations were important and managers
pressed forward with the use of temps as a source of cost savings, but the manner in which they were used and rules governing their work were significantly shaped by institutional factors.

Due Process and Equal Opportunity

Although little remarked on by NPE scholars, one of the most widespread shifts in ILMs—a shift certainly as important and as widespread as the diffusion of performance-based pay—has been the implementation of a broad range of personnel practices associated with equal employment opportunity and with non-union due process and grievance procedures. These practices obviously have important implications for career pathways and promotions within firms.

A large literature has emerged seeking to explain the origins of these practices within firms. Much of it is rooted in the “institutional” stream of thought in organizational sociology, which focuses on how organizations adopt practices to legitimize themselves in the eyes of other powerful actors or because the practices enhance the power of interest groups within the organization (Selznick 1949; DiMaggio and Powell 1983). The latter perspective is clearly congruent with the argument I have been developing here. The challenge is to distinguish the institutional explanations from more efficiency-based arguments claiming that the adopted practices make economic sense for the organization.

Although it is not always possible to prove definitively that efficiency considerations are unimportant, the body of research taken as a whole does convincingly demonstrate that institutional considerations have played a central role in the diffusion of the new practices and in determining the specific forms that they have taken. In an early foray into this field, Edelman (1990) studied the adoption of grievance procedures in non-union firms, specifically by sampling 52 firms in the San Francisco Bay Area. Her broad argument was that organizations adopted these practices not because of any specific legal mandate (there was none), nor because of unions (the firms were non-union and a variable measuring whether they faced a union threat was insignificant). Rather, it appeared that the general normative environment coming out of the Civil Rights movement placed a heavy moral or societal weight on firms to build “fairness” into their procedures, and firms with personnel departments were more likely to respond to this pressure.

In later research Edelman, Abraham, and Erlanger (1992) placed even more emphasis on the role of personnel departments in pushing for practices that were self-aggrandizing. Specifically, they studied firms’ adoption of practices related to wrongful dismissal charges and showed that the personnel profession as a whole systematically exaggerated the risks that firms faced in these law suits, in some cases using estimates that were “absurd” (p. 65) and in other cases simply misleading (by, for example, reporting data from California as if it applied to all states). As result, senior managers were convinced to implement personnel practices that enhanced the power and role of the personnel professionals.

In yet another set of studies, Dobbin et al. (1993) and Dobbin and Sutton (1998) examined the adoption of promotion policies in response to Equal Employment Opportunity mandates from the Federal Government. There is certainly an underlying efficiency explanation in that firms faced the risks of lawsuits and contract loss if they did not respond in some way. However, underlying the specific practices that emerged are important subtleties. First, Dobbin et al. showed that the particular practices adopted converged on a set of personnel practices that were isomorphic with the procedurally oriented, quasi-judicial administrative configuration of the federal government—formal, merit-based, employment and promotion conventions complete with an internal system of grievance adjudication. (p. 402)

In other words, the particular practices that were adopted were based upon an imitation of those of the more powerful external organization, not upon an internal optimization logic. Of course, an NPE response to this might point to the search and information
savings that such an approach might bring. Dobbin and Sutton (1998), however, demonstrated that even after federal pressures sharply diminished in the Reagan years, and hence the costs of non-compliance fell sharply, personnel professionals continued to press for new practices, and they shifted their rhetoric from that of legal risk to the alleged internal benefits of diversity. Put differently, the personnel staff constructed (without any evidence) a new argument once the legal risk argument was no longer valid. A fair reading of this literature is not to suggest that efficiency explanations were unimportant or irrelevant; rather, it is to say that they are strikingly incomplete and perhaps even naive with respect to the forces that drove the adoption of the wide range of personnel practices associated with equal employment opportunity and fairness in the workplace. The self-interest of a specific group within the firm, as it competed for resources and power, was a big part of the story.

Conclusion

In this paper, I have aimed to describe an institutional perspective on ILMs that recognizes the role of standard efficiency and market considerations yet at the same time brings to bear a deeper understanding of how organizations make decisions about organizing work. One central characteristic of my argument is that it does not develop a universal model parallel to the universalistic tendencies of neo-classical economic theory. Rather, my argument is pitched at the level of the firm or organization. Nor does it address the psychology of individuals (whether they are utility maximizers or not) or the operation of external markets. The creation of employment rules within organizations is the terrain on which I think that institutional ideas have the most purchase.

It is important in this context to understand that even at the level of the organization, the approach proposed here does not reject market forces, for this would be foolish indeed and particularly so in a period when established forms of work organization are under a great deal of pressure from competition. Rather, my goal has been to develop a more sophisticated understanding of the dynamics of Internal Labor Markets. The central proposition, one which has a long tradition in the literature, is that these systems are the result of a political process in which groups with competing objectives and norms play out a contest. Market or efficiency considerations are certainly one player in this contest, and in the examples I have presented I have made an effort to give them their due. But they are not the only player. It is also important to recognize that the resources available to the different players in this contest will vary over time. It would not be unreasonable, for example, to argue that in the period spanning the Depression and World War II, the external environment favored internal actors who opposed NPE-style thinking whereas today, cost-minimization advocates hold an advantage. But this ebb and flow does not alter the fundamental argument.

At the same time that market forces are given their due, it is also important to understand that the institutional considerations emphasized here are persistent and are not simply transitory aberrations that will be quickly eliminated by the force of competition. First, the market has proven to be very slow in extinguishing long-standing practices, as the history of the automobile industry makes clear. In addition, important streams of research and theory in economic models of technology adoption, the sociology of organizations, and political science models of institutional change all argue that the forces of path dependence are strong and that practices become locked in.

The emphasis on group processes and politics, however, points to an important limitation of this paper: the question, not

---

8 I am grateful to Sandy Jacoby for clarifying this point for me.

9 See, for example, Nelson and Winter (1982); Arthur (1989); Hannan and Freeman (1984); Stinchcomb (1965); and Thelen (2004).
addressed here, of how groups are formed and re-reformed. This, of course, goes to the issue of how people and collectivities come to identify themselves as having commonalities and how they frame what they regard as legitimate interests and goals. Although this is a topic of growing interest (Barley 1989; Scully and Segal 2002; Piore and Safford 2006), it would take this paper too far afield to enter into it.

The goal of good social science should be to understand the determinants of the empirical realities that we observe, both at a point in time and as they change. Elegant theory is not an end in itself. This is important to understand because the arguments developed in this paper are certainly less clean and elegant than are NPE models. Moreover, these arguments do run the risk of devolving into what has been a weakness of institutional theory, namely a tendency simply to provide a long list of factors that should be taken into account, without any strategy for assigning relative importance or for testing the arguments. To say that institutional theory runs this risk, however, is not to say that it should be abandoned. Rather, advocates of institutional models have to be willing to be clear about their arguments and about the strategies they use for testing them. The gains from such an effort should be a richer understanding of how employment systems are evolving and what drives their evolution.

REFERENCES


