ADMINISTERING LABOR CONTRACTS USING TRANSACTION-COSTS ECONOMICS

KENNETH VOGEL
Loyola Law School

Labor contracts are rules of governance between workers and employers over time. The efficiency of the transaction and the relationship can be affected by the provisions of the contract. It is argued that the characteristics of the firm and the workers determine what kinds of contracts and governance of the provisions of contracts are most efficient. Among other results, it is argued that collective bargaining, together with grievance procedures and arbitration, is the most efficient form of labor contracting and governance for large firms with skilled work forces.

Professor Oliver Williamson, who has emphasized the importance of considering transaction costs in analyzing how and why firms are organized,[1] has recently extended transactions costs analysis to modes of governance of contractual relationships (Williamson, 1975: 233). Based on the costs of acquiring information important to that transaction, he develops a scheme for characterizing the governance of transactions as belonging logically to one of four modes: market, trilateral governance (arbitration), bilateral governance (negotiation), or legal integration (internal organization). One of the most important and widely studied areas of contract law is the labor contract. It affords an excellent opportunity for examining a wide range of types of governance that depend upon the information required in the transaction and therefore should allow us to apply Williamson's schema to deepen our understanding of some of the reasons for the diversity of contracts.

LAW & POLICY QUARTERLY, Vol. 5 No. 1, January 1983 129-152
© 1983 Sage Publications, Inc.

In this article I will first provide a summary of the literature on transaction costs as they apply to the possible forms of organizing productive relationships: the market, the integrated firm, or the complex contract. The second section will examine the special nature of labor as a factor of production. Finally, I will inquire into the nature of labor agreements and the parties to them (firms, workers, and unions) to see if transaction costs economics can help to explain the features of agreements and some of the reasons that other types of governance do not appear.

I. TRANSACTION COSTS

Whenever people desire to make an agreement, whether it be for economic gain, personal satisfaction, or even purely altruistic reasons, there are costs expended to make that agreement. These costs can usually be broken into three categories: costs of obtaining information, costs due to uncertainty regarding the future, and the increased costs of the actual bargaining process in the context of a small number of actors. Clearly there are a whole range of transactions having these costs in varying degrees. The model of a pure market as used in neoclassical economics is an example of a type of transaction having (essentially) zero transaction costs. This market transaction is assumed to take place in a world with perfect information, perfect foresight, and large numbers of actors all offering the good at the same price. Although those are strong assumptions, they seem to do a good job of explaining markets for many types of goods, particularly those of essentially homogeneous quality sold at known prices for immediate delivery. As we drift away from these assumptions to conditions where goods are meant to be delivered in the future, where quality can vary, when prices are not well known and may need to be set in the process of negotiation, transaction costs become more significant.

Transaction costs are significant in the context of this article when they are high enough that it may be efficient to organize transactions in a different form from the pure neoclassical

economic market. The most extreme departure from the market form of organization is the integration of transacting parties into a single firm. This vertical integration of buyer and seller will obviously be most efficient when there are likely to be recurring transactions between the parties. But that is not enough to explain integration. There are after all costs of organizing a firm. Even as market transactions are not costless, neither is it costless for a hierarchy to direct the actions of a supplier and a user of intermediate products. The primary cost to such vertically integrated firms is lack of choice of suppliers. Once a firm has purchased its supplier, it no longer will find it as easy to purchase in the market when it is dissatisfied with its supplier divisions.

After all, the advantage the market gives is choice among a large number of suppliers. When one is unsatisfied with a supplier, that supplier can be dropped and another one substituted. Once legally integrated, the costs of dropping a supplier are internal; no longer can the threat of loss of revenue be used to discipline a supplier that is part of the integrated firm. The discipline in the integrated firm must be hierarchical; either the supplier division must be directed to change its practices so that its product and price are satisfactory, or the division chief must be replaced with one who will come closer to satisfying the purchasing division. Neither of these options is costless to the firm, as each involves a change in production techniques. Or, as Coase has commented in a general way: "a firm will tend to expand until the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organizing in another firm" (Coase, 1937: 285).

One may then ask when it is most likely to be more efficient to organize a transaction internally. As the market is most efficient when the traded goods are homogeneous and are provided by many suppliers, internal organization will be most efficient under the opposite set of conditions. For example, when a firm has a need for a particular good (which, for example, must be custom-made) and the supply market consists of a small number of firms, the supplying firm may be able to garner a large markup,

especially if it is able to learn how to make this customized good more cheaply than its competitors, and the information on this production process is not available to other potential suppliers.[2] A purchasing firm in this monopolist-monopsonist relationship may feel that it should share in that surplus and may choose to merge to get it.

Of course, there is a position these two parties may take that is intermediate between market and merger. The parties may choose to define their relationship in terms of a long-run contract. Often two parties may not wish to make a complete merger of their operations for all time[3] merely to lower the costs of some market transactions. There are two common situations where parties may wish to develop a more complex contract than is assumed by strict neoclassical economics. The first situation occurs when two parties contract for one to fabricate a one-time customized good for the other party, and it will take time between the agreement and delivery to produce the good. The second situation is when the buying party expects to continue to need a noncustom good for repeated delivery in the future and where both parties know the extent to which the supplying party can enjoy decreasing costs by being assured of a market for a certain amount of its production.

In both of these cases the two parties will want some protection against changes in the relationship that increase their costs. For example, suppliers may want to be assured that they have a chance to adjust and recoup their investment if buyers, either because of changes in circumstances (in the macro economy) or new opportunities, or change of mind, decide to end the relationship. They may also want some assurance that, if their costs change in some unforeseen way, the entire risk does not fall on them alone (so that they might be able to raise their prices). Buyers, on the other hand, want to be assured that the quality of the product does not vary downward. They also want to be assured that the quantity and timing meet their needs, and that the price, if it varies, does not change unexpectedly or unreasonably.

Unfortunately, as people are not privy to information about what will happen in the future and as they are not always able to

account for every circumstance, and therefore are unable to write contracts dealing with all possible circumstances, the parties may be forced to depend in the future on nonmarket, nonhierarchical procedures to adjust their differences. Sometimes they will agree on machinery in their contract to help adjust to unforeseen circumstances;[4] otherwise, they will depend on negotiation, if there is goodwill, or legal action, if there is not.

When labor is the commodity to be supplied, purchasing firms are able to use only two of these forms of governance: market or contracting. In the United States, at least, the permanent sale of labor is forbidden as slavery, so merger is not a legally possible form of organization. In order that we might inquire into the nature of labor contracts and their governance, in the next section we shall consider the nature of labor as a traded commodity and explore some of the reasons why labor is different from other commodities and other inputs.

II. LABOR MARKETS

It is crucial to keep in perspective the basic differences between labor and other commodities. A sale of labor entails an agreement between workers and employers that the workers give up their freedom over the use of their time and their bodies for a given payment (see Simon, 1957). A worker is then a supplier of an input to a productive process, but the worker continues to have ultimate control over that input even after the process has begun. Unlike suppliers of raw material or capital goods, workers can continue to vary the quality or quantity of the labor input, even after they have entered into the control of the employing firm. It is not feasible to separate the input—labor—from the supplier—the worker—both technically and psychologically, and this fact contributes to the enormous complexity of the administration and management of the employing organization.

Employers would like to pay no more than the marginal product of a worker. To the extent that marginal product declines as labor supplied increases, there is a surplus the employee would also like to share. There are a number of ways that employees can

be paid closer to their maximum wage (their average product). The most obvious is to have the employer agree to pay close to their average product. When that is not done, workers have an incentive to shirk; that is, they can reduce their work effort so that they will be paid close to their actual average product. It really does not matter if the worker is paid a salary, an hourly wage, or a piecework rate. There are alternative schemes for shirking available to each. When paid a salary, workers may simply try to work fewer hours, or they may use methods available to wage and piece-rate workers. Workers paid a wage, for example, can attempt to produce less, and workers on a piece-rate can attempt to produce lower quality pieces.

It is in the interest of employers to devise methods of supervision, and rules regarding employee behavior to try to counter the shirking of employees. The closer employees can get to receiving their maximum possible wage, the greater the incentive on the part of the employers to develop supervisory procedures, as their profit is minimized in this case. In fact, we should expect employers to expand supervision to the point where the increase in the revenue product of workers equals the marginal cost of supervision (see Alchian and Demsetz, 1972). It is important to remember that there will always be an informational asymmetry between workers and employers. Workers know how much better they could work, among other things; while employers know the details of the supervisory and other procedures and the likelihood that an employee may be disciplined.

The most extreme discipline an employer can use is termination.[5] Termination can be a costly procedure to adopt. Discharging an experienced employee can lead to a real cost to the degree that an employee with some experience at the firm has learned to perform tasks more effectively for a given wage than would a new employee.[6] The more heterogeneous the supply of labor for any position is, the more costly replacing a worker can be. Of course, termination can be very costly to the employee as well. Employees who are fired lose their wages (less any unemployment benefits) and will have the fact of having been fired as

part of their record when they seek work elsewhere. This information may cause potential employers to lower their estimates of the worker's potential productivity.

The extent of the use of any discipline is determined by the same internal profit-maximizing calculus that employers use for all their decisions. Employers should take disciplinary actions only if the marginal decrease in direct productivity is less than the decrease in costs due to shirking. One of the reasons we find that disciplinary actions will not always be efficient is that it is costly to get the information to have an effective disciplinary system.

A less disciplinary management style, which is prevalent in Japan for example, attempts to elicit cooperation from workers by assuring them that many types of discipline, including layoffs, will not be used in normal circumstances (see Noda, 1979).[7] Cooperation often elicits information from workers, which may lead to lower costs as well as decreases in many types of shirking. Their familiarity with the equipment which they operate often enables workers to design more appropriate techniques for its use, or the workers may even be able to modify their equipment to increase its productivity. This kind of cooperation is less likely without incentives to assure workers a fair share of the long-run gains.

Unfortunately, it is difficult for an employer to gather information on either the extent of shirking, or on the extra effort expended by an employee. Obviously, the employee will wish to underreport the former and overreport the latter. By supervising, an employer should receive more direct and relatively less-biased information than is likely to be supplied by a worker, However, as the size of the organization increases, the employer will find it difficult to engage in direct supervision. Having a large staff of employees, therefore, will either dilute the amount of supervision or it will necessitate many supervisors and possibly many layers of supervisors. Not all supervisors will carry out company policies in the same way, and supervisors are inherently more difficult to oversee than lower-level workers with structured jobs or ones in which performance can more readily be measured. While it may be possible to tell when workers in this latter category are not

performing up to their expected standards, apparent malperformance by supervisors may be due not to their unwillingness or inability to follow approved techniques, but rather to unobservable differences in the work force.

Supervisors therefore have much discretion in what they do. As management of supervisory personnel is more troublesome than supervision of production workers, it is easier for supervisors to perform their duties in a manner contrary to the policies of the firm. This will be especially true in a strictly hierarchical organization where information on supervisory behavior comes either from direct management observation or from information supplied directly by the supervisor. Supervisors, like other workers, have incentives to present their best case to their managers. Without useful information on output, which is unreliable insofar as it may be due to exogenous differences in work forces, managers have weak sources of information regarding supervisory performance, through the normal hierarchical structure of the firm.

Supervisors can therefore take advantage of their position for personal reasons to reward or punish workers under them. They may also misunderstand the prescribed procedures of the organization and unwittingly contradict organizational policies, or they may believe that the firm's policies are ill-advised and consciously try to confute those policies. Often nonhierarchical procedures may be necessary to allow workers to complain about misapplication of organizational policy to the supervisor's superior. Protection of workers and the possibility of eventual reward for enforcement of firm policy are important to encourage this action when appropriate.

Rather than dealing with authority-relationship problems, firms may wish to simplify their hierarchical structure and use the market as an alternative to an authority relationship when dealing with labor. Firms have three alternatives to the direct hiring of labor. For an economist, the most obvious choice is to substitute capital for labor. Alternatively, firms may choose to substitute whole departments or divisions and go into the market for intermediate inputs rather than producing them internally.

Finally, a firm may choose to hire subcontractors to fulfill parts of the firm's functions. In subcontracting situations, all the labor performing those functions is under the control of the subcontractor, but the timing of the function and the provision of (some of) the capital for the subcontractor is controlled by the hiring firm.

Once again, the choice of the extent to which a firm will use internal labor versus its substitutes will be determined by the relative ratio of marginal product to marginal cost of each. The only departure from simple economic theory is that now the internal costs of hierarchical organization and the relevant contracting and transactions costs of using the market are explicitly included in the costs of the alternatives.

If complex contracting procedures can help to lower costs in normal input contracts, they should be able to do the same in contracts for labor. The next section explores various contracting procedures, from individual labor contracts to collective bargaining contracts, to see if transaction costs economics can help to explain and predict why certain provisions should be found in contracts under different conditions.

III. LABOR CONTRACTS

There are essentially three issues that must be dealt with in any contract: a description of what is being traded, a price, and a mechanism for making adjustments when there are disagreements as to whether the terms of the contract have been carried out. If the parties do not agree specifically on these issues, contracts can occasionally be saved by implying what the parties must have meant.[8] In labor contracts the three vital matters are a job description, including duties, promotion, seniority, and so on; and a wage and a grievance procedure. Often, in very simple contracts, the first and the third matters are implied, not agreed upon. Management usually reserves the right to direct the employee to do anything that is not illegal. The hours of work are also determined by the employer, but are frequently limited by

law. The grievance procedure can be very informal—simply a complaint to the boss—or very formal—a legal proceeding in court.

Determination of the fair or efficient wage is the aspect of labor contracts most often considered by economists. There are two dominant theories in wage determination: human capital—a microeconomic theory based on the characteristics of the worker, and market determination—a macroeconomic theory based on the characteristics of the economy. Others, of course, consider the noneconomic theory linking union power to the relative level of the wage in unionized versus nonunionized industries (see Parsley, 1980). Wage determination is basically not a transaction cost problem and will not be dealt with here.[9] Rather, differences in labor contracts in the areas of job description and grievance procedure will be analyzed using relevant transaction costs.

A. Work Rules

When labor is considered to be a homogeneous input, employers will wish to have complete discretion over the tasks and conditions of work. They will seek to assign workers to tasks and provide working conditions, such as hours worked, time off, and physical surroundings, so that profit is maximized. Workers will have little power to complain in this case. As labor is homogeneous, any worker can be replaced by any other. If there is an excess supply of workers, replacements will be available. Only to the extent that there is a range of profit maximizing conditions used by employers, so that some employment is easier or more desirable, will there be any differences in wages. This case with homogeneous labor corresponds most closely to the conditions needed for a neoclassical market.

Examples of this kind of employment relation are not uncommon even in modern industrial societies. Migrant farm workers, emergency snow shovelers, and common laborers are examples that come to mind. They are workers with little human capital and low opportunity cost of their time who often may not even be

eligible for unemployment compensation. Efficiency in these markets means that all monopoly gains, to the extent that they exist, go to the employer.

However, for most modern employment relations, a division of labor is much more common. Efficiency due to the division of labor can occur for a number of reasons. Workers can train themselves in particular skills needed for different types of employment. Professionals like physicians, attorneys, or school-teachers are examples of specially trained workers, as are many white-collar employees: bookkeepers, secretaries, and computer operators, to name but three. Also, many skilled blue-collar jobs require, or are made easier by, previous training: automobile mechanics, electricians, heavy equipment operators, and the like. Workers in these occupations are equally useful to a number of potential employers who desire their special skills.

Some workers are formally trained by their employers. Factory-authorized repairmen, management trainees, and airline pilots fall into this category. In addition, most workers develop their skills more extensively by performing the tasks required of their jobs. In these two cases, workers develop or improve their human capital in such a way that while they have increased their value in their job category, they have increased it even more for their present employer. Clearly the form of the employment agreement will vary for these different categories of workers. Workers will have more incentives to be productive if they know something about their expected duties and about the employer's willingness to refrain from exploiting all the rent of their labor skills.

Let us first consider the likely characteristics of contracts for the group of workers who learn to improve their performance of tasks by actually performing them on the job. When employers know that workers will be more efficient if they perform only one task, employers will, even in the absence of an agreement with the workers, refrain from exercising their power to have a worker perform many unrelated tasks. Workers may not attempt to become as proficient as possible unless they have reason to believe that they will benefit personally from their increased productivity

and that they will not have additional or different tasks assigned to them. Therefore, it will be in the employer's interest (as well as the employee's) that certain protections be given to employees.

What are the nature of these protections giving employees the proper incentives even in the presence of the transaction cost due to lack of information on the actual abilities of employees? Employees will want to know what tasks are expected of them. They will want to know that if they are hired as, say, machine operators, they will be expected to operate and perhaps maintain their machines, but that general maintenance is not part of their job descriptions. This assurance has two objectives: it allows workers to concentrate on the tasks they know that they need to learn, and it may provide them with some self-esteem by eliminating some tasks from the range of duties expected of them.

Employees also expect some pledge that they are expected to work a certain number of hours. If they must work more hours, they will be compensated and if they must work fewer hours, there must be a valid reason for this decrease and a fair method of deciding who will continue to work. The seniority system has been developed in part to deal with this last concern. The employees who are most valuable to an employer are those who are most productive. If human capital increases with time on the job, those workers who have been there the shortest time will tend to be least productive. A seniority system, by decreasing time worked for the least senior employee first, tends to leave the most productive workers on the job. Employers would like to differentiate, if they could, between more productive junior employees and less productive senior employees, and they would wish to keep those more productive junior employees. But due to the transaction cost of information of the worker's productivity, they may not be able to choose the proper worker at low cost. Therefore, it may be more efficient, on average, to use a seniority system, even without considering the effects on worker security.

Considering the problem of having many supervisors with different standards and different personalities, it should be clear that if employers keep their discretionary power to lay off (or transfer, or promote, or demote, or give preference as to working

hours or extra training) on the basis of ability, the supervisors will be the ones to have the discretionary power. In large organizations the ultimate managers may not want to allow too much discretionary power to rest in the hands of supervisors or foremen who may be inaccurate or biased in their assessments. They may also wish to reward loyalty and provide workers with the assurance that they will not be treated arbitrarily. Plainly, the larger and more complex the organization, the more likely formal rules will be substituted for discretion when considering who is laid off, rehired, or given overtime. The information costs to the ultimate manager are much lower in a small, simple organization, and cost minimization (including information costs) is more likely to be consistent with managerial discretion.

As to the other two groups of employees, with human capital due to formal training, the assurances likely to increase the self-respect and productivity of the informally trained worker are just as important. Indeed, because these workers have more human capital, it is easier for them to find alternative employment. An employer whose cost of replacing workers increases with either increased or more specific human capital will be willing to give stronger asurances to those employees.

How do employees and employers negotiate in order to achieve the requisite meeting of the minds for a contract to exist? There are essentially three ways that such contracts can be formed. Employers can offer employment, with whatever assurances they think are most efficient, and potential employees can either accept and be employed or reject and not be employed. Employers can also negotiate separate contracts with each employee. Finally, employers can negotiate one contract with an agent of a group of employees—a union.

Let us first consider separate negotiation with each employee. When the organization is very small and the difference between employees is very large, this system may have some purpose in the negotiation of work rules and salaries. However, efficient as it may be to have a particularized wage schedule in small firms to allow discretion in awarding differential wages based on ability (which is more cheaply and accurately estimated than in large

firms), to negotiate work rules (conditions of employment) separately for each employee is bound to cause havoc. First, any individual negotiation procedure has very high direct transaction costs from negotiation itself. The more negotiations that must take place, the higher the cost of negotiating. Large organizations, therefore, would have very high costs if they conducted such a procedure, and have incentives to find procedures that lower their costs.

Aside from that obvious concern, if work rules differ for similar employees, it will be obvious that, and how, they differ. Whereas differences in wages (including fringe benefits) might be kept confidential, differences in tasks or working conditions are easily discernible by all affected parties (there are low costs for workers to acquire that type of information). These differences are bound to cause resentment. Workers who feel they have been discriminated against will not have the same incentives to work up to their capability, especially when there are not clear bureaucratic rules encouraging workers to be more productive to qualify for benefits.

The negotiation technique of the employer offering a detailed contract on a "take-it-or-leave-it" basis obviously has its benefits for employers. They can supply the terms they think are most efficient. They can always structure the employment opportunities to be bureaucratized, so that they can eliminate the disadvantage of seemingly arbitrary differences in the treatment of employees. They can be sure that they meet the opportunity costs of the employees whenever they willingly take a position with their firm. However, this technique ignores an important potential (transaction) cost saving—no nonmarket information from employees is considered in developing a contractual package of work rules and wages, simply because there is no negotiation.

Internal efficiency for the firm is not equivalent to efficiency of a transaction. Whereas internal efficiency in hiring requires a firm to equate (given) marginal costs with marginal revenue product, efficiency in a market requires the marginal benefits of both

parties to be equal. If marginal costs (prices) are not given, there may be many internally optimal allocations. If parties to a transaction can discuss tradeoffs among wages and benefits and work rules and working conditions, they can potentially exploit all the gains from the transaction. Not only can employees provide information on which issues are important to them, but they may often be able to provide information on which set of work rules, for example, which division of labor, is most efficient. In summation, take-it-or-leave-it negotiation is likely to work best when this last piece of information is known to employers (most likely in a small firm, with lower costs of direct supervision), when the costs of negotiation become larger than the savings from the information gained in negotiation (most likely in a large firm with unorganized employees and hence high costs of negotiation), or when the opportunity costs of employees are small (e.g., in company towns, or in periods of high unemployment).

In spite of the potential for high transaction costs (due to strikes, etc.), contract negotiations with a union afford an opportunity to develop a more efficient contract through negotiation. In large firms, when the transaction costs of individual negotiations are high, a single negotiation for at least a class of workers provides a means for minimizing those transaction costs while providing information on worker preferences to increase the efficiency of the contract. That potential gain aside, the primary increased transaction cost with union negotiation (apart from the social and private costs from strikes) is the problem of reconciling differences in preferences among workers.

The more heterogeneous the work force within a union, the more likely it is that differences in preferences as to the terms will be important. If the work force is of dissimilar age and experience, the older and younger members will differ on the importance of a seniority system versus a system of preferences based on perceived ability. The more homogeneous the union is as to human capital and required tasks, the easier it will be to agree on work rules, promotion standards, and the like. There are

essentially two ways to attempt to reconcile these differences between workers: internally in a single union or by separate unions based upon employment position.

Both of these organizational forms are found in unions. Industrial unions typically take the first form. All workers in an industry (or all industrial workers in a firm) negotiate one contract. These unions are most efficient as negotiators when differences between workers in the union on the basis of human capital are small. With small differences in human capital, seniority can form the basis[10] for classification into positions, as well as the basis for wages and choice of hours worked. When there are large skill differences, craft unions are more common. The issues important to differently skilled workers may vary greatly when the type of work is very heterogeneous. For example, in the building trades, electricians and carpenters employ different tools to practice different techniques. Their knowledge of skills should make separate negotiation of work rules cheaper.

The difference in transaction costs due to the two union types is due to the difference in costs between hierarchy and negotiation as forms of reaching agreement. A craft union should have lower internal organizational costs than an integrated union made up of workers with diverse skills. However, the cost of negotiations become higher as the number of negotiating parties increases, just as in the case of commercial bargaining. Other than the lowered number of bargainers, which lowers bargaining costs, as compared with private negotiation, it should be clear that the differences between different crafts and different work rules need not appear as arbitrary as they would with private (individual) negotiation. The choice between forms of union organization (on efficiency grounds) ought to be based on the relative costs of bargaining and hierarchy, and bargaining agents should diversify to the extent that decreases in the marginal cost internally gathering information and resolving disputes among members outweigh the increases in the cost of bargaining between unions and employers. There is unfortunately no mechanism to insure this result, for while all internal costs of organization fall on the union,[11] some of the increased costs of bargaining fall on the

employers and they are prohibited from influencing the form of union organization.[12]

After the contract is negotiated, its provisions are not always perfectly clear, or they may not be applied consistently by different supervisors. Some method of contract governance must be used to interpret the contract during its term (which in the United States is usually three years).[13] In the next section we consider the three forms of contract governance: market, negotiation, and arbitration; and we consider the conditions under which each is appropriate as a transaction cost minimizing procedure.

B. Governance

(1) Market

Market governance implies that if the provisions of a sale are unsatisfactory, a different supplier or a different purchaser will be substituted for future sales. In the context of a labor agreement, firms will fire unsatisfactory workers and workers will quit if the actions of the boss are unsatisfactory. These responses are nearly always available as a last resort; however, the instances when they are efficient as the only available mechanism are much rarer. Contracts providing only a market mechanism (or resort to court proceedings) make the enforcement of any detailed provisions very expensive.

The costs to the employer and to the employee of using the market mechanism have been mentioned above. They are lowest for employers when labor is homogeneous and unemployment is high; they are lowest for employees when there are many alternative opportunities for employment. They do not vary with the importance of the issue, so only very important differences, for example, failure to perform job tasks to a minimal standard, will be resolved in this matter. The cost of courts also does not vary much with the importance of the issue. These costs include time waiting for resolution, direct legal fees, and the possibility that the publicity resulting from the disagreement might discourage others from dealing with the parties in the future. Market

mechanisms are therefore very clumsy for dealing with contracts with job descriptions, a specified system of promotion, choice of hours or order of layoff, or other provisions that may enhance the efficiency of the agreement.

(2) Negotiation

The bilateral governance procedure allows the parties to meet after the specific facts of a possible contract violation (or merely an area of uncertainty as to contract interpretation) are known, and to come to a mutually satisfactory resolution, if that is possible.[14] The normal use of a bilateral governance (grievance) procedure in a labor contract will be to appeal to a higher management official the actions of a particular supervisor in his or her decision to discipline or to promote, contrary to the intention of the contract. The process may be single- or multi-stage; that is, it may be possible to make further appeals of the decision at increasingly higher levels of management. This procedure gives management, in a complex hierarchical organization, a low-cost method of ascertaining when low-level supervisors are failing to follow the wishes of upper management. If the limitations to which management has agreed do in fact increase productivity, it is important to the firm that they be followed.[15]

> The rules which have been developed with regard to selection, layoff and retention, promotion, and discipline and discharge have resulted in significant limitations upon the arbitrary exercise of managerial prerogatives and power. These limitations are not simply the result of trade union pressures through collective bargaining; they are more in the nature of *self*-restraint which managements have imposed upon themselves as a result of organization needs for coordination, specialization, and personnel regulation (Vollmer, 1960: 17).

The procedure is universally initiated by an aggrieved employee as management, in the absence of specific contractual forbearance, has the power to do as it wishes. Employees will

either carry out the process on their own, or the union will provide representation. It is very inefficient for employees to bear the responsibility and cost of the grievance process on their own. First, there are skills in negotiation that may be acquired through practice, and it would be inefficient for all employees to invest in their acquisition. Second, it is often intimidating for employees to complain to their bosses that they have been treated unfairly; with no effective outside enforcement of provisions prohibiting employer retaliation, there may be a chilling effect on the bringing of complaints (so that fewer than the efficient level would be brought). Third, there may be economies of scale in negotiation. Fourth, a union, with limited resources, will have an incentive to bring only cases that are conceivably meritorious, or at least there is an incentive to limit the number of stages to which a meritless grievance will be taken. Finally, with constant negotiating between partners who are fairly evenly matched, a system of precedents in contract interpretation is sensible. This alone will reduce nonmeritorious claims and will encourage consistency of treatment, from which the efficiency advantage of rules is derived.

Collective bargaining agreements have been called "an effort to erect a system of industrial self-government" (United Steelworkers v. Warrior and Gulf Navigation Company, 1960: 580) which generates a "new common law—the common law of a particular industry or a particular shop" (United Steelworkers, 1960: 579). The development of the common law takes place between contract negotiations when the statutory law of the industry or shop is amended. As the contract negotiations take place between union and company, they are the parties affected by the agreement, and it is most efficient for their preferences to be of primary importance in the grievance procedure. However, the union gets its legitimacy from the consent of its members, and it has a duty, as their agent, to represent all members fairly.

The duty of fair representation has been much discussed in the legal literature (see Cox, 1956; Feller, 1973; Summers, 1977). It is generally agreed that when the collective agreement gives the union exclusive control, in order to enhance efficiency in the formation of the common law of the contract over the grievance

procedure, the individual employee can sue the employer for breach of his individual rights only after first showing that the union has acted unfairly in failing to process the grievance through all the stages of the procedure (Vaca v. Sipes, 1967). The main requirement is that the union "protect equally" all those members that it represents (Steele v. Louisville and Nashville Railroad, 1944). Equality of representation is also a transaction cost minimizing rule.

If the union can discriminate between grievants on the basis of the facts of their case, equality of protection implies that only in those cases in which the union acted arbitrarily would the duty of fair representation be breached. If the union determines that further grieving is fruitless, it has the power to stop, so although employees are assured that they will be treated in a nonarbitrary manner, which is the point of the complex contractual rules and procedures, the union will be able to grieve effectively those cases it believes it has a chance of winning.

With more complex and complete agreements, the procedure for negotiating logically becomes more efficient with collective (union) control. When agreements are simple, specifying only wage scale and hours but no protection or work rules, there is no collective interest in contract interpretation, and therefore, there are no collective preferences to be followed as to the relative importance of competing claims. This type of agreement is presumed to be enforced by individual employee's suits for breach (Cox, 1956: 605). However, when contracts are more complex, it is consistency itself that encourages efficiency.

(3) Arbitration

Arbitration is a means of eliminating the high cost of formal legal procedures or market mechanisms (quitting or striking) when the parties are unable to agree. It has been called "the means of solving the unforeseeable by molding a system of common law" (United Steelworkers, 1960: 581). As the importance to the parties of maintaining the relationship increases, it becomes more

efficient to use a system of arbitration as a last resort. Arbitration allows the parties to choose the third party who will decide the merits of the grievance. It enables arbitrators to become experts in the narrow area of a particular contract (especially if the contract has, or will have, over the long run, a large number of grievances brought to arbitration). Under these circumstances arbitration is clearly more efficient that the alternative dispute-resolution techniques. Arbitration is essentially universal in union contracts: 99% of collective agreements include some sort of grievance machinery and 98% of those industrial agreements include a provision for submitting grievances to an impartial arbitrator when the ultimate management and union representatives are unable to agree (U.S. Department of Labor, 1974: 64). Yet in 1962, while 54% of nonunion firms had formal grievance procedures, none of them appeared to involve arbitration as a final step (Steele and Fisher, 1964: 265).

Without the more complex agreement typical of union contracts, there does not appear to be the same interest in consistent nonarbitrary interpretation of the terms of the agreement. Nonunion firms probably are willing to trade off lower productivity from their workers, due to the unsureness of nonarbitrary treatment, for lower wages. For the firm this allocation may be equally efficient as compared to the complex union contract, but there are potentials for gains among the workers. These gains will not be realized if their expected capitalized value is lower than the expected costs of forming and maintaining a union.

IV. CONCLUSION

Efficiency is now understood to include the minimization of the technical costs of production and the transaction costs of contracting. The efficiency of any arrangement depends upon tradeoffs made among all the costs. There are a few simple rules that ought to be followed in the transaction cost area of labor agreements to make the most efficient tradeoff between various

types of transaction costs. We have argued that efficiency in the system of labor agreements depends on two factors: the degree of hierarchy in the firm and the human capital of workers.

Larger, more hierarchical firms need a mechanism to review the day-to-day decisions of lower-level supervisors. It is therefore relatively more efficient for them to provide such a mechanism in their labor agreements to ensure the participation of line workers in the review process. So grievance procedures are more likely in large firms.

Firms with workers who have more specific training (human capital) and/or better alternative employment opportunities will have greater costs from losing those workers, These workers will demand, successfully, to be assured that they will be treated in a nonarbitrary fashion. Therefore the more such workers a firm has, the more likely it is to have complex work rules and/or a seniority system.

Finally, once a firm has a complex agreement, for one of the reasons stated above, the enforcement of the agreement of bilateral or trilateral governance is less costly than if strikes or court action are needed to settle disputes. The issues tend to become more subtle and intricate in disputes regarding complex contracts so they are both more likely and more difficult to resolve. Parties involved in a long-term agreement can then jointly pursue those matters which are most central to them and to the maintenance of their efficient relationship.

NOTES

1. This obviously follows from the seminal work of Ronald Coase (1937). Williamson's work in the field is most clearly developed in Williamson (1975).

2. Under these circumstances the supplier firm is said by Williamson to have a first mover advantage.

3. Mergers of course need not be for all time, but even if firm A, having purchased firm B, decides later to dispose of those assets, there is no reason to presume that the original owner of firm B would be the likely purchaser. Nor is there any reason to presume what the price would be.

4. Well-trained attorneys will advise their clients to include some sort of grievance or adjustment procedure in their long-term contracts.

5. In the Netherlands, parties to private employment relationships have been prohibited since 1945 from unilaterally dissolving such arrangements. The economics of this special arrangement is analyzed in Martin (1977).

6. Such an employee is said to have specific human capital. (See Becker, 1964.)

7. Noda quotes three haiku to describe the difference between Japanese and Western management: "If the cuckoo won't sing, kill it," which advocates the firing of unproductive workers; the carrot and stick approach: "If the cuckoo won't sing, make it sing"; and the modern Japanese style, which assumes that people find work natural: "If the cuckoo won't sing, let's wait until it does."

8. For example, the Uniform Commercial Code provides that under certain conditions, market price will be used when the price is left out in a contract for the sale of goods (*U.C.C.* s2-305).

9. Some recent work has been done on the effects of uncertainty on wages using both a macroeconomic model, Hall and Lilien (1979) and a microeconomic model, Lazear and Rosen (1979).

10. A study of 400 union contracts in effect in 1970 found that 92% had some seniority provision, and 10% of those applied other factors, such as ability (*Labor Relations Expediter* at 479, 1970).

11. Of course, the larger the union is the more money it can raise from dues and the more leverage it is likely to have, at least in politics. This benefit must also be weighed against increases in internal organization costs.

12. "It shall be an unfair labor practice for an employer . . . to dominate or interfere with the formation or administration of any labor organization or contribute financial or other support to it," 29 U.S.C. s158(a) (2).

13. In 1973, 51% of workers covered by 1339 major agreements had three-year contracts (U.S. Department of Labor, 1974).

14. If it is not possible and the issue is important enough, the parties may still resort to market or formal legal procedures.

15. "My experience [is] . . . that most initial collective bargaining agreements contain little more than rules governing employer conduct that correspond to existing employer policy or practice," (Feller, 1973).

CASES

STEELE v. LOUISVILLE AND NASHVILLE RAILROAD (1944) 323 U.S. 192.

UNITED STEELWORKERS v. WARRIOR & GULF NAVIGATION COMPANY (1960) 363 U.S. 574.

VACA v. SIPES (1967) 386 U.S. 171.

REFERENCES

ALCHIAN, A. and H. DEMSETZ (1972) "Production, information costs, and economic organization." Amer. Economic Rev. 62: 777-792.

BECKER, G. S. (1964) Human Capital; A Theoretical and Empirical Analysis With Special Reference To Education. New York: Columbia Univ. Press.

COASE, R. H. (1937) "The nature of the firm." Economica 4: 386-393.

COX, A. (1956) "Rights under a labor agreement." Harvard Law Rev. 69: 601-621.

FELLER, D. E. (1973) "A general theory of the collective bargaining agreement." California Law Rev. 61: 665-678.

HALL, R. and D. M. LILIEN (1979) "Efficient wage bargains under uncertain supply and demand." Amer. Economic Rev. 69: 868-890.

LAZEAR, E. and S. ROSEN (1979) "Rank-order tournaments as optimum labor contracts." National Bureau of Economic Research, Paper 401.

MARTIN, D. (1977) "The economics of employment termination rights." J. of Law and Economics 20: 187-198.

NODA, M. (1979) "Business management in Japan." Technology Rev. 81: 20-31.

PARSLEY, C. J. (1980) "Labor union effects on wage gains: a survey of recent literature." J. of Economic Literature 18: 1-24.

SIMON, H. A. (1957) Models of Man: Social and Rational. New York: John Wiley.

STEELE, J. and P. FISHER (1964) "Effects of unionism in southern plants." Monthly Labor Law Rev. 87: 258-279.

STIGLER, G. and K. BOULDER [eds.] (1975) Readings in Price Theory. Chicago: Irwin.

SUMMERS, C. (1977) "The individual employee's rights under the collective agreement: what constitutes fair representation?" Univ. of Pennsylvania Law Rev. 126: 251-265.

U.S. Department of Labor (1974) "Characteristics of agreements covering 1000 workers or more." Washington, DC: Bureau of Labor Statistics.

VOLLMER, H. M. (1960) Employee Rights And The Employment Relationship. Berkeley: Univ. of California Press.

WILLIAMSON, O. E. (1975) Markets And Hierarchies: Analysis and Antitrust Implications. New York: Free Press.

KENNETH VOGEL is currently Acting Professor of Law at Loyola Law School, Los Angeles. At the time the article was written, he was Assistant Professor of Economics at the State University of New York at Buffalo. He has a Ph.D. in Public Policy Analysis and a J.D. from the University of Pennsylvania and does research in the effects of law and regulation on individual choice.