UC Berkeley

Other Recent Work

Title

Bargaining and Influence Costs and the Organization of Economic Activity

Permalink

https://escholarship.org/uc/item/32s7d4jv

Authors

Milgrom, Paul Roberts, John

Publication Date

1987-02-23

Peer reviewed

UNIVERSITY OF CALIFORNIA, BERKELEY

Department of Economics

Berkeley, California 94720

Working Paper 8731

BARGAINING AND INFLUENCE COSTS AND THE ORGANIZATION OF ECONOMIC ACTIVITY

Paul Milgrom and John Roberts

February 23, 1987

Key words: Transaction costs, contracting, bargaining theory, economic systems, organization theory, rent seeking

Abstract

A series of efficient short-term contracts can support an efficient long-term relationship, even when the parties to the contracts behave opportunistically and specialized investments are needed. However, specialized assets, private information, and costs of measuring quality can all raise short-term bargaining costs, resulting in losses from the decentralized, market approach.

Centralized governance involves assigning discretionary rights to intervene or to resolve disputes to a central office executive. Increasing centralization raises costs in several ways, most notably by raising the returns to politicking and other influence activities. If one defines a firm to be the smallest business unit largely free of outside discretionary intervention, then mergers of firms increase influence costs. Influence costs are also incurred in some market transactions as well as in the public and non-profit sectors.

JEL Classification: 511, 026, 053

ACKNOWLEDGMENT

We thank Patrick Bolton and seminar participants at the Hoover Institution for their helpful comments. We also thank the National Science Foundation and the Guggenheim Foundation for their generous financial support.

Bargaining and Influence Costs and the Organization of Economic Activity

by Paul Milgrom and John Roberts

Until recently, economists had paid little attention to the internal workings of business organizations. In the abstract models that form the core of microeconomics, a firm is simply a collection of possible production plans together with a rule for selecting among them. Typical rules include profit or share value maximization for firms run by the owners of capital or average wage maximization for labor-managed firms. In these highly abstract models, the processes by which firms generate and evaluate decision alternatives, coordinate distant branch stores, factories, and offices, balance the competing interests of employees, owners, customers, suppliers and creditors and motivate them to work in the general interest, though not excluded, are not explicitly represented.

¹It is sometimes argued that these core economic models are inconsistent with the ideas (1) that firms are managed by individuals with a limited ability to process information, calculate and make consistent decisions and (2) that firms respond to changing circumstances in the short run simply by following their established procedures or routines. To evaluate this argument, notice that the formal theory, as presented for example by Debreu (1959). incorporates the possibilities that a production plan can be uncertain or contingent, using inputs and producing outputs in a way that depends on emergent events. Such production plans can include uncertain R&D, organizational routines, the use of boundedly rational supervisory and managerial personnel as inputs to production, the development of skilled managers as outputs, etc. Consequently, the propositions established by Debreu and others in the context of abstract, general economic models necessarily apply also to more specific and detailed models in which firms may be actively managed by boundedly rational managers or in which a firm's short run behavior is determined by "standard procedures," provided that in the "long run" firms choose their plans according to the specified rule.

Why have economists clung for so long to such an incomplete account of business organizations? Historically, the chief task of economic theory has been to explain how market economies, with so little centralized direction, could have performed as well as they have. Economic growth in the West, where market economies originated, has been rapid and steady over a very long period of time. Shortages (characterized by rationing and queues) have been rare. Jobs have been found for a diverse and growing workforce with its ever-changing mix of education, skills, and experiences. On those infrequent occasions when firms have produced goods that nobody wants or excessive amounts of goods that people buy in limited quantities, the problem has typically been short-lived. The great economic accomplishment of the West is not that its firms, carefully managed, could achieve order in their affairs, but that its markets, amidst all the chaos of variety, new products and processes, diverse suppliers, and growing and shifting populations have, with no apparent planning, directed the available resources to such good effect.

History seems to be confirming in Japan and the newly industrializing countries that Western economic growth was no accident. In Japan, since the Occupation Forces imposed a capitalist democracy on the country after World War II, economic growth has been explosive. The post-war economic development of Korea provides similar evidence. In 1986, per capita GNP was \$2,270 in capitalist South Korea

and \$800 in communist North Korea, and the difference appears still to be growing. $^{\pm}$

Since Adam Smith's original explication of how markets might quide economic activity to serve the public interest, economists have dissected, analyzed, refined, and formalized the theory of markets controlled by an impersonal force -- the "invisible hand." But even as the economists worked, vast changes were taking place in Western economies. No longer were firms mostly family affairs with bookkeeping and management operations done at night, when the shop was closed. The nature of business changed. Continuous production processes and specialized equipment came into use and production activities came to be controlled by the very visible hands of engineers, chemists, and professional managers. In the United States, as Chandler (1977) has explained, the growth of the railroads and the telegraph opened national markets, made it economical to produce in large scale factories that strained the capacity of local suppliers, and so required more careful planning by factory managers. With planning came both the opportunity and the need to consider explicitly alternative ways of organizing production. Should an automaker make or buy its headlamps, batteries, and body parts? Should it own a network of dealers selling directly to the public, contract with existing distribution companies, or sell to independently owned retailers? Purchasing supplies and hiring services in the market came to represent just one organizational

As reported in the "Business Bulletin" of the Wall Street Journal, Western edition, January 15, 1987, page 1.

alternative for acquiring the necessary inputs for production, assembly, distribution, and sales.

What determines which inputs a firm will acquire by ordinary market exchange and which it will produce itself? What difference does it make whether a firm produces an input for itself, has a regular supplier produce it, or buys it on the market from the lowest bidder? The second formulation of the question shifts attention subtly away from the mechanical details of how production is arranged toward a focus on how the relationships between the people who carry out the successive stages of production are managed. It suggests that whether production is arranged internally or externally need not determine what equipment will be used or which people will do the work; "internal" and "external" production are just terms to describe in a very partial way how a productive relationship is to be managed.

Economists who study organizational questions have come to see the market as but one alternative for solving the management problem of coordinating the diverse activities and interests of consumers and firms. Markets can then be fairly evaluated only in comparison to other means of solving the same problem. Such an evaluation cannot be made until a unified theory of management processes has been developed. Without such a theory, economists' recommendations about such bread-and-butter economic questions as whether to regulate a monopoly and whether public or private organizations should provide services like education, communication, transporta-

tion, etc. must be regard as tentative, at best. The Economics of Management is a subject that economists can no longer ignore.

Currently, the dominant approach to analyzing the organization of economic activity is transaction costs economics. This approach replaces the emphasis on technology and the management of production of earlier economic theories with and emphasis on transactions and the management of relationships. In section 1, we articulate the transaction costs economics approach. We criticize transaction costs theory in section 2 for its failure to consider production costs and transaction costs together and for its failure to identify properly the transaction costs that are characteristic of markets and those characteristic of alternative modes of organization. We argue that the crucial transaction costs that are distinctive of markets are those associated with negotiating short-term agreements and that the corresponding critical costs of internal organization are those that arise from the presence of centralized discretionary authority. Among the latter, political or influence costs (the losses that are suffered when individuals seek to influence the organization's decisions in order to advance their private interests and when the organization adapts to control this behavior) are probably the most pervasive. Section 3 is devoted to a study of bargaining costs and section 4 to the costs of centralized authority. In section 5, we apply our analysis of the costs of centralized authority to investigate the costs of government regulation and intervention in the economy and of the uncertainties in the law that increase the discretionary authority of judges and juries. Our general conclusions are summarized in section 6.

1. Transaction Costs Economics

By making that shift in perspective away from the mechanics of production and toward a focus on production relationships, Coase (1937) created transaction costs economics. According to the dictionary, "transact" as an intransitive verb means "to do business with; negotiate." "Transaction costs" encompass the costs of deciding, planning, arranging and negotiating the actions to be taken and terms of exchange when two or more parties do business; the costs of changing plans, renegotiating terms and resolving disputes as changing circumstances may require; and the costs of ensuring that the parties perform as planned or agreed. Transaction costs also include any losses suffered on account of inefficient group decisions, plans, arrangements or agreements, inefficient responses to changing circumstances, and imperfect enforcement of agreements.

The central hypothesis of transaction costs economics is that in capitalist economies, production will tend to be organized so as to economize on transaction costs. Thus, in particular, inputs will tend to be acquired in the market rather than produced by the firm when the costs of market transactions are less than those of internal transactions. The hypothesis does not specify how this

Partie American Heritage Dictionary of the English Language. Boston: Houghton Mifflin Co, 1980.

tendency to economize on transaction costs arises. Careful planning by especially competent management may sometimes be responsible*, as may imitation of successful firms by less successful ones or the growth of firms that organize their affairs efficiently and the collapse of firms that fail to do so.

Coase's hypothesis as specified above is too vague to be confronted directly with evidence. To develop specific predictions from the theory, it is necessary to identify the costs associated with transacting business in different ways and to discover how circumstances cause these costs to vary.

Oliver Williamson (1985) has proposed one framework within which Coase's theory can be made more specific and operational. Williamson's theory is based on an analysis of the costs of contracting in business relationships. Contracts govern a firm's relationships with its suppliers, employees, customers, creditors, and shareholders. A central premise in Williamson's theory (foreshadowed in Coase's own work) is that any contract that calls for the future delivery of a good or service, the future provision of capital, or the future performance of work must inevitably be incomplete. There are several reasons for this. First, the parties cannot perfectly anticipate all of the many possible contingencies that may affect their costs of performing as promised, or even their

[&]quot;Chandler's (1962) impressive account of how the managements at DuPont, General Motors, Sears, and Standard Oil of New Jersey discovered the multidivisional structure indicates the extent to which some managers can innovate to rationalize the organization of production.

[&]quot;See Nelson and Winter (1982).

ability to perform as promised. Second, even for those circumstances that can be anticipated, it is often more economical to decide how to respond when the need arises rather than to plan in advance for every foreseeable contingency. The third reason derives from a fact often emphasized in twentieth century philosophy: Language derives its meaning from context and shared experiences, and so the precise limits of application of any sentence or description, especially as applied to novel future circumstances, must be indeterminate. For that reason, drawing contracts with too many fine distinctions increases the likelihood that emerging events will fall into an area of actual or potential ambiguity, leading to disagreements which will have to be resolved after—the—fact.

What are the consequences of this incompleteness of contracts? If planning and contracting were complete and costless, the parties to a contract would, after their initial agreement, be led to act as one. They would determine in advance and in detail the best possible actions for every contingency that might arise and the contract would specify that those actions be taken. In reality, because planning and contracting consume real resources and because perfectly explicit and freely enforceable contracts cannot be written, the parties content themselves with an agreement that frames their relationship, that is, one that fixes both general performance expectations and a set of institutions to govern deci-

^{*}Lindblom (1959) elaborates the value of postponing decisions and not attempting always to optimize.

⁷See Quine (1960).

sionmaking in situations where the contract is not explicit and to adjudicate disputes when they arise. The differences among simple market contracting, complex contracting, vertical integration, and other ways of organizing transactions lie primarily in the institutions they specify for governing the relationship when circumstances not foreseen in the contract arise.

For the transaction costs theory to explain the great variety of contracting practices that actually exist, it must identify the critical dimensions that favor one form of contracting over another. According to Williamson: "The principal dimensions with respect to which transactions differ are asset specificity, uncertainty, and frequency. The first is the most important and most distinguishes transaction cost economics from other treatments of economic organization, but the other two play significant roles."

Asset specificity refers to the degree to which the value of an asset depends on the continuation of a particular relationship. For example, a firm that rents a computer system from its manufacturer and invests in software and training for its employees to use the system has invested in specialized assets (if an identical or perfectly compatible computer cannot be rented or purchased from another source) — the software and employee training would lose much of their value if the firm switched to another computer. A supplier who acquires specialized dies or locates a plant or warehouse near his customer's remotely sited factory has similarly invested in specialized assets. Klein, Crawford, and Alchian (1978)

^eWilliamson (1985), page 52.

have dubbed the flow of profits that the investor stands to lose from terminating a particular business relationship "appropriable quasi-rents." Logically, appropriable quasi-rents exist precisely when there are specialized assets.10

For concreteness, let us suppose that a supplier invests in specialized assets. The supplier's worry is that his customer might behave opportunistically, that is, he might seek to force a reduction in future prices, or he might curtail his purchases, make unreasonable quality demands, increase the variability of demand and the number of rush orders, or take other actions that diminish the supplier's margins. If the assets were not specialized, these threats would not be great cause for concern: The supplier would be protected by his option to shift the assets to other uses where they

In the language of economists, "rent" to a supplier consists of that portion of the price of a good (or service) in excess of the price needed to attract the supplier to provide the good. Together, rents plus "quasi-rents" refer that portion of the price beyond the amount necessary to prevent an active supplier from ceasing to produce. Positive quasi-rents exist when the current price is sufficient to keep an existing supplier active but not sufficient to attract new suppliers into the industry.

Normally, quasi-rents arise as a return on a sunk investment. For example, an employee whose high pay is attributable partly to long hours invested in learning the ins and outs of his company is said to be earning quasi-rents.

As we use the terms here, buyers can also earn rents or quasirents. These can be measured as the difference between the prices that buyers pay for their inputs and the highest prices they would be willing to pay.

An asset is an owned factor of production, something that contributes to the production of goods or services. To say that it is specialized means that its value is greatest in one particular relationship. But another way to describe the same situation is to say that the relationship itself is an asset, a factor that increases production. It is a specialized asset because its value — the "appropriable quasi-rents" — evaporates when the relationship is terminated.

command an equal return. However, by definition, specialized assets cannot be shifted to other uses without loss, so the investor may be forced to accept reduced margins leading to a substandard return on his investment.

Indeed, it has frequently been argued that concerns that the buyer will appropriate quasi-rents may lead the supplier to invest too little in specialized assets, or to demand a high price initial-ly to compensate for the risks involved. 11 Klein, Crawford, and Alchian cite the case of Fisher Body as an illustration. In the 1920s, it refused to build its plants adjacent to the General Motors plants that it served. They argue that Fisher Body properly feared that such a plant siting would make it vulnerable to subsequent attempts by GM to force a reduction in its margins.

One apparent option to mitigate this problem of "appropriation of quasi-rents" is to make the price and other terms of the contract more explicit and rigid and to impose greater penalties for breach of contract. However, this solution is itself costly. To the extent that the number of clauses in the contract remains unchanged, additional rigidity means reducing the parties' flexibility to respond to future circumstances. If clauses are added in an attempt to specify in advance more contingencies and a correspondingly greater variety of responses, direct contracting costs are increased and the problem of ambiguity in the language of the contract becomes

¹¹Variations of this argument are given by Williamson (1986), Klein, Crawford and Alchian (1978), Grout (1984), Tirole (1986), and Grossman and Hart (1986). We will examine this argument critically below. In what follows, we recap the argument as it has been made before.

greater: There is a greater likelihood that the actual circumstances will fall into a fuzzy region where it is difficult to identify which clause should govern.

It is primarily the presence of <u>uncertainty</u> about what circumstances will prevail when future actions must be taken that makes complete contracting impossible. Greater uncertainty about what actions will be appropriate make rigid contracts more likely to lead to bad decisions, and therefore more costly. Flexible contracts, too, entail costs; they do little to reduce the risk that quasiments will be appropriated. In this context, Coase's hypothesis is that the parties will normally agree on the kind of contractual arrangements in which these costs are least.

If the opportunity to appropriate quasi-rents arises <u>frequently</u> in a particular supply relationship, the parties may find it economical to craft a specialized governance structure to deal with it. Depending on the nature of the transaction, there may be many alternative governance structures available which may vary greatly in their complexity and costs. At one extreme in terms of simplicity are short, flexible, generally worded contracts, to be interpreted by the courts in the event of a dispute, but where the parties rely primarily on each others' goodwill, business reputation, and standard procedures and on their continuing business relationship to smooth out disagreements without extensive bargaining. For many purposes, especially for arrangements to deliver standard commodities at an agreed price, simple contracts may be entirely adequate. In other situations, more careful planning or

governance may be needed. Then, contracts can be made more detailed, for example by including price escalator clauses and clauses indicating the penalties for breach of contract or how to deal with specified contingencies; is a procedure for selecting and using arbitrators or private judges can be substituted for courtroom litigation; or firms can merge and assign authority for decisions to an executive. Highly detailed contracts and specialized decision or dispute resolution procedures are expensive to write or design, but the costs of writing and designing are fixed costs which, once sunk, can be applied again and again to similar transactions. Hence detailed contracts and specialized procedures are most cost-effective when similar transactions are conducted with great frequency.

The general predictions of what Williamson calls the "governance branch" of transaction costs economics can be summarized as
follows: In comparing business relationships that function in the
same legal environment and at the same point in time, governance
structures will be most complex and finely crafted for those
transactions with (1) the greatest value of appropriable quasirents, (2) the greatest uncertainty surrounding the conditions of
performance, and (3) the greatest frequency of transaction.

Before turning to our critique of the theoretical developments just described, we digress to emphasize and explicate an aspect of

¹²Levi Strauss's television advertising contract for the 1980 Olympic games, signed in 1977, gives some idea of how detailed (and prescient) some contracts can be. It specified that in the event the American team did not compete in the Games, no payments would be due.

THAS GM and Fisher Body did in the example cited above.

the fransaction costs theory that many readers will find puzzling. Coase's conjecture holds that affairs will be arranged in the way that minimizes transaction costs, regardless of the bargaining power of the contracting parties. This hypothesis might seem to require that the party with the bargaining power is the one who incurs the transaction costs: Why would the party with all the bargaining power economize on the other party's costs? For example, consider an employer-employee relationship in which the employer holds "all the bargaining power." By this we mean that any net gain from the relationship beyond what the two parties could have earned on their own accrues to the employer. Why would anyone expect the employer economize on the employee's share of transaction costs? The answer is that if the employer does so, he can reduce the employee's wage by the amount of the cost savings. 14 Similarly, if the employee had all the bargaining power and simply rented assets from the capitalist for a fixed fee, the employee could negotiate a lower fee by economizing on the capitalist's costs.

In effect, Coase's hypothesis is that the parties to a transaction negotiate over governance arrangements to reach an efficient

^{**}The same sort of argument is commonly made to support the proposition that firms, regardless of labor bargaining power, act to minimize production costs. Both arguments are subject to the same sorts of objections. For example, if the employer must provide costly general training to the employee to minimize costs and if the employee is unable to pay for the training and unable to bond his performance, the employer may be unable to capture the gains from the training and so may make the inefficient decision to provide no training.

service, quality control, and rapid delivery. Thus, by alleviating customers' concerns, it enhances demand for the product. Thus, competition can be regarded as a substitute for subtle governance structures. Indeed, if one views the problem of governance as one of specifying an appropriate role for third parties, competition might be regarded as a governance device in its own right.¹⁷

Whatever its merits in reducing governance costs, the practice of licensing a chip to a competitor results in a loss of production efficiency. There are economies of scale in chip production as well as economies of experience ("learning curve" effects) that are lost when a second firm is involved in the production. Moreover, the licensor may incur substantial costs just to transfer its chipmaking technology to the licensee; the semiconductor industry literature is replete with examples of second sources who took a year or more (which is a long time in the fast moving semiconductor business) before they could successfully produce the chip they had been licensed to make, despite extensive assistance from the licensing firm.

To recapitulate, the chipmaker's problem involves simultaneously choosing who will produce the chips and how the supplier-customer relationship will be managed. The decision involves weighing both manufacturing costs and transaction costs. Moreover, considering these costs together leads naturally to the somewhat

¹⁷Farrell and Gallini (1987) have offered a similar analysis of second sourcing, and have generalized beyond the licensing application to consider other ways in which a monopolist might encourage competition in order to better govern its relationship with its customers.

surprising conclusion that competition itself can be usefully regarded as a governance device. Thus, evaluating organizational arrangements generally requires that attention be paid to <u>all</u> costs.

Our second point of criticism of the theory described in section 1 is that the theory fails to identify the relevant transaction costs. We will show that the costs of negotiating suitably detailed short-term contracts are the key to evaluating the efficacy of transactions organized through markets. If the costs of negotiating short-term contracts were always zero then, we claim, organizing economic activity through market exchange would always be perfectly efficient. On the other hand, when the costs of negotiating periodic exchange agreements are sufficiently high then, regardless of such other factors as the presence or absence of specialized assets, there are important savings to be realized by placing the activity under centralized control, so that potentially costly disputes can be settled quickly by the central authority.

To understand the meaning of these propositions, one must first understand what we mean by the terms "short-term" and "bargaining costs." When describing contracts, short-term refers to a period short enough that all relevant information for decisions to be made during the period is already available. Short-term contracts can be complete, because there is no need to specify how to act in various

reflere, we echo Coase (1937), who held that the transaction costs of the price system consist of the costs "of discovering what the relevant prices are" and "of negotiating and concluding a separate contract for each exchange transaction" (pages 391-2). We include both of these categories in our expansive definition of "bargaining costs."

contingencies. We interpret the term "bargaining costs" expansively, just as we did the term "transaction costs," to include all the
costs associated with bilateral bargaining, competitive bidding, and
other voluntary mechanisms for determining a mutually acceptable
agreement. Bargaining costs include not only the wage paid to the
bargainers: or the opportunity cost of their time, but also the
costs of monitoring and enforcing the agreement and any losses due
to failures to reach the most efficient agreement possible.

With these definitions, zero short-term bargaining costs means that the bargainers require negligible physical and human resources to reach efficient short-term contracts. (A short-term contract is efficient if there is no other feasible short-term contract that both parties would prefer to sign.) However, the bargainers cannot commit themselves in any short-term contract to restrict their long-term behavior. For example, the parties may be able to agree on what investments in specialized assets to make this year and who will pay for those investments, but they cannot commit themselves to behave benignly next year toward the party who, having paid for the investment, finds himself with appropriable quasi-rents.

To establish the key role of bargaining costs, suppose that the costs of negotiating short-term contracts were zero. We consider a supplier-customer relationship in which the relevant parties are each risk neutral, financially unconstrained, expected-wealth-

¹⁹⁰r to those who prepare bids and those who solicit and evaluate them in case competitive bidding is used to set the price.

maximizing bargainers and for which efficient production demands that the supplier, the customer, or both invest in specialized assets. The parties are assumed to be both rational and opportunistic in the sense that their current behavior does not depend on past unbonded promises or on how past costs and benefits have been distributed. Finally, we assume that it is prohibitively costly to write contracts that govern prices and behavior in the distant future because there are too many contingencies that would need to be evaluated and described (that is, too much uncertainty), but that it is costless to write contracts that govern the prices, bonus payments and actions to be taken in a near future period for which the relevant conditions are already known.

In general terms, the situation is as follows: There is opportunistic behavior, imperfect long-term contracting, specialized assets, and uncertainty ruling the future. According to the arguments advanced by Williamson and by Klein, Crawford, and

gainers maximize the expected present value of profits at some fixed discount rates or at discount rates that properly reflect the correlation of project returns with other aggregates risks. These objectives for the firm are widely used in the theory of financial economics.

Interestingly, the assumption that the parties are risk neutral expected wealth maximizers without financial constraints plays an important role in our analysis. The assumed absence of financial constraints severely limits the applicability of the following analysis to the problem of investments in human capital, because laws against slavery prohibit the use of human capital as loan collateral. The assumption is also likely to fail in applications to public projects involving health, safety, or environmental quality - projects for which the public's preferences are not easily expressed in risk neutral expected pecuniary gain terms. We make no claims here about the importance of bargaining costs in assessing the relative efficiency of market contracting for these applications.

Alchian, these conditions are sufficient to prevent a market arrangement based on a series of short term contracts from yielding an efficient outcome. Nevertheless, we claim that if bargaining costs were zero — a condition that is apparently consistent with our other specifications — then the market outcome would be efficient. That is, the actions taken by the parties both in the short—run and in the long—run would in all contingencies be identical to those that would have been specified in the "ideal contract" — the efficient (possibly long—term, complete) contract that the parties would sign if there were no restrictions at all on contracting. 21

We will explain the argument supporting this proposition and the defect in the received theory below. Before doing so, we wish to emphasize two points about the proposition. First, as in the labor contracting example discussed in section 1, the actions taken under an efficient contract do not depend on the bargaining power of the parties. Only the distribution of the fruits of the bargain depend on bargaining power. Conversely, because the parties are risk neutral, if the actions they take coincide with those that would be specified in the ideal contract, then the arrangement is efficient, regardless of the stream of payments made between the

arThis proposition, which will be stated and proved formally in another paper, was inspired by the related propositions of Crawford (1986) and Fudenberg, Holmstrom, and Milgrom (1987).

EMActually, this conclusion requires that there be no "wealth effects" on the parties' preferences and that the parties have unlimited access to investment funds. These conditions are implied by our assumption that the parties are risk neutral expected profit maximizers.

parties. Second, we do not claim that the parties' inability to write complete contracts leaves their bargaining power unaffected or that it has no effect on the way risks are shared. For example, an ideal contract might have specified that if the technology used in one party's specialized investment becomes obsolete, the second party will reimburse the first half of the unamortized portion of the cost of the investment, according to some specified amortization schedule. With short-term contracting, such an agreement would most often be impossible, so the distribution of risks is affected by the restriction to short-term contracting. What is unaffected is the set of actions that the parties will eventually take, and hence the efficiency of the agreed arrangement.

We will not provide a mathematical statement and proof of our proposition here. Instead, we illustrate the proof in the context of a relevant example.

Consider the relationship between Fisher Body (the supplier) and General Motors (the customer) analyzed by Klein, Crawford, and Alchian. (We reviewed their analysis in section 1.) Suppose, initially, that the relationship lasts for two periods. In the first period, the parties reach an agreement about plant siting and design (investments in specialized assets) and how the cost of constructing the plant will be divided between the parties. Such an agreement does not require knowledge about future contingencies and specifies only the immediate actions to be taken and how they will be compensated. In the second period, the parties negotiate prices, possibly a fixed transfer payment, quality standards, and a delivery

schedule in full knowledge of the circumstances then prevailing (e.g., current model year body designs, demands for the various models of autos, costs and availability of steel and substitute materials, etc). By our assumption of costless bargaining, regardless of the first period agreement, the second period agreement will be efficient given the then prevailing conditions (including the results of decisions taken in the first round).

Now consider what would happen if the parties were to agree in the initial period to make the efficient siting and plant design decision. Then, the actions taken in the second period would, in all circumstances, agree with those specified under the hypothetical ideal contract. We therefore conclude that there exists a short-term contract that could be signed in the first period that would lead to efficient decisions being taken in both the first and second periods. Actually, by varying who pays for the initial investment in the plant, all distributions of the fruits of these efficient decisions can be attained. So, any contract that leads to inefficient decision making can be improved upon for both parties by some contract that leads to efficient decision making. Therefore, if the costs of short-term bargaining were zero, then the agreement

prevail, so the design decision they reach cannot be based on that information. By the efficient decision, we mean the decision that would be agreed to on the basis of the available information if perfect long-term contracting were possible.

andur assumption of opportunism implies that the future agreements reached and actions taken do not depend on how the cost of the initial plant investment is divided.

reached would indeed be one that leads to efficient actions. This argument can be extended to encompass any number of dates of transaction.

What then, was wrong with the argument advanced in section 1? Why shouldn't the fear of opportunism by General Motors make Fisher Body unwilling to enter into the arrangement? The answer is that Fisher can be compensated for the risk by having General Motors bear part of the cost of the plant. Why, then, shouldn't General Motors fear that its quasi-rents will be appropriated by Fisher? Because the agreement can call for General Motors to pay for only as much of the plant's earnings as it expects to appropriate in future negotiations. Threats of appropriation are simply distributional threats; they are not threats to efficient action as long as bargaining costs are zero. Distributional threats among risk neutral parties with common beliefs and no private information can be compensated by an initial cash payment. The efficiency of market arrangements are limited only by the costs of negotiating efficient short-term contracts. This conclusion points to the central importance of bargaining costs in determining the efficiency of market transactions. We shall study the origins and determinants of bargaining costs in section 3.

^{**}Grossman and Hart (1986) reach the opposite conclusion in their model of bargaining with specialized investments, despite an apparent assumption of zero bargaining costs. There are two key differences that distinguish their model from ours. First, they do not assume -- as we do -- that the bargainers are risk neutral expected wealth maximizers. Second, they assume that even short-term contracts cannot be written: The parties cannot contract about what investments to make in specialized assets and how to divide the investment costs.

Our third criticism of transactions cost theories concerns their relative silence regarding the source, nature and magnitude of the transaction costs incurred in non-market transactions. Indeed, despite the firm beliefs of many economists that markets often hold great advantages over non-market forms of organization, the absence of these factors in the transaction costs theory leaves it unclear (at least in terms of this theory) why market modes of transacting are ever to be preferred to non-market ones.

Identifying the costs of general non-market transactions is a task to be approached with great caution. As Chandler (1962) has documented, business organizations have changed substantially and repeatedly over the past century, and the disabilities (transaction costs) suffered by an older form of organization may be overcome by its replacement. Nevertheless, in section 4, we shall argue that there are generally identifiable costs of internal organization. A "firm" is distinguished in our analysis by its substitution of some amount of centralized authority for the relatively unfettered negotiations that characterize market transactions. And, we argue, there are certain transaction costs that inevitably accompany the existence of centralized authority.

markets that he holds that "organizations are a means of achieving the benefits of collective action in situations in which the price system fails"! Here, he interprets organizations broadly. "Formal organizations, firms, labor unions, universities, or government, are not the only kind. Ethical codes and the market system itself are to be interpreted as organizations."

3. Bargaining Costs

What are the costs of bargaining? We have defined these above to include the opportunity costs of the bargainers' time, the costs of monitoring and enforcing the agreement, and any costly delays or failures to reach agreement when efficient production requires that the parties cooperate. Our analysis in this section will focus on costly delays and failures to reach agreement.

The idea comes easily to economists that when parties to a bargaining situation have all the relevant information, they will agree to some efficient bargain. Nash (1950, 1953) elevated this proposition to an axiom in deriving his famous "bargaining solution." Coase (1960) made it the linchpin of his theory of property rights. Buchanan and Tullock made the same point in connection with their argument that only costs of private bargaining can justify government provision of goods or services:

We shall argue that, if the costs of organizing decisions voluntarily should be zero, <u>all</u> externalities would be eliminated by voluntary private behavior of individuals regardless of the initial structure of property rights. There would, in this case, be no rational basis for state or collective action beyond the initial minimum delineation of the power of individual disposition over resources.

The evidence supporting this idea, however, is mixed. When experimental subjects are asked to divide a fixed sum of money, say \$10, they have little difficulty agreeing to split the sum equally without costly delays or disagreements. But when the thing to be

PBuchanan and Tullock (1962), pp. 47-48. Emphasis in original.

amSee Roth (1986), for example.

divided is more complicated - so that symmetry does not focus the bargainers attention on an obvious solution - posturing, haggling, and disagreement is more likely to occur, as each party seeks to create or stake out a reasonable-sounding position that gives it a large share of the available rewards.

To get a better idea of how serious these difficulties of coordination might be, we turn to the analysis of a <u>barqaining by demands game</u>, introduced originally by John Nash (1950, 1953).

Thus, suppose that two parties have \$1 to divide. The rules are as follows. Each of the two parties, A and B, makes a demand, a and b. If the demands are consistent with the available resources, that is if a + b does not exceed \$1, then each party gets what it has demanded. If the demands are inconsistent, both parties get a payoff of zero.

If the problem were presented in just this way, the parties would very likely each demand \$.50, resulting in a 50-50 split. In the terms used by Thomas Schelling (1960), the 50-50 split is an obvious focal point — a way for the parties to coordinate their demands. Most real bargaining situations are complex, and focal points may be absent or, what is just as bad, there may be many focal points. What should we expect to happen then?

For a game theoretic analysis, we may ask: What is the full set of non-cooperative equilibrium outcomes of this demand game?

These outcomes represent patterns of behavior that are consistent with rational and well-informed self-interest seeking on both sides.

The answer is that for any pair of positive numbers summing to \$1 or

less, there is a Nash equilibrium (possibly in mixed strategies of the demand game at which the players' payoffs are precisely those numbers. There is also a Nash equilibrium in which both bargainers demand the whole \$1 and, as a result, both receive zero. Thus, there is no compelling reason to assume that the bargainers can or will always succeed in coordinating their demands to allow an efficient agreement to be reached, even when both parties understand the circumstances and possibilities perfectly well. Quite the contrary: "Rational" bargainers can squander an arbitrarily large fraction of the available resources trying to obtain favorable bargains for their respective sides.

Remarkably, the presence of markets with alternative suppliers and customers bidding for business virtually eliminates the potential for inefficiencies in bargaining. Suppose, for example, that the bargaining situation is one that involves two suppliers and a buyer. In terms of our model, there are three parties to the bargaining - A, B, and C, who make demands a, b, and c, and the demands are compatible if either a + b or a + c is less than \$1.

The probabilities adopted by a player at equilibrium represent the other player's subjective uncertainty about the demand his counterpart will make. The existence of mixed strategy equilibria means that the players can consistently be uncertain about what demands their counterpart will actually make without attributing irrationality or mistaken views to the other player.

The Actually there are many such equilibria, but to verify the claim it suffices to exhibit one. Let x and y be positive numbers summing to one or less. Suppose that player 1 demands x with probability y/(1-x) and demands 1-y with the complementary probability, and that player 2 demands y with probability x/(1-y) and 1-x with the complementary probability. This describes a Nash equilibrium in which player 1's expected payoff is x and player 2's is y.

The rules of the game are as follows. If the demands of the buyer are inconsistent with both suppliers' offers, no agreement is reached and each party receives a payoff of zero. Otherwise, the buyer A does business with the supplier making the smaller demand, or randomizes if the demands are equal. If the buyer and a seller make consistent demands, each of them receives the amount he has demanded. Almost all the "equilibria" of this "auction" version of the demand game are efficient. The Moreover, the buyer receives all the surplus at equilibrium; this is just as in the competitive market outcome.

Variations on this three-party demand game lead to the same conclusion. For example, suppose that if the demands are consistent, one or the other party gets \$1 minus the other party's demands, or that the parties split the difference. In each of these games, essentially all of the equilibria lead to the efficient outcome in which all the surplus accrues to the buyer. This is the natural result of bidding competition among the suppliers.

The demand games just described can be interpreted as models of a competitive supply market. When perfectly competitive suppliers

based on the idea of iterated dominance, applied to this game with an infinity of strategies for each player. To compute our equilibrium, we first limit the parties to name integer multiples of some discrete unit, e.g. pennies. Next, we eliminate weakly dominated strategies to create a new game with a smaller strategy set and iterate, applying the dominance criterion recursively until no further reduction in the set of strategies is achieved. We then compute the Nash equilibria of the reduced game. At every such equilibrium, each supplier demands approximately \$0 and the buyer demands approximately \$1. Passing to the limit as the unit of account grows small isolates this equilibrium of the original game.

must make simultaneous offers, competition among them reduces the scope for disagreement with the buyer, leading to efficient outcomes. The two-party demand game, by contrast, illustrates the inefficiencies that may result when there is but a single supplier. Specialized assets tend to give rise to a sole supplier condition, which is accompanied by a struggle for rents and consequent bargaining inefficiencies. Thus, specialized assets <u>cause</u> bargaining costs, which may explain the predictive successes of the received transaction costs theory.

The first class of bargaining costs, then, take the form of "coordination failures"; they arise when there exist multiple patterns of mutually consistent self-interested behavior that individuals could adopt, and market institutions fail to ensure that only the efficient patterns actually emerge. Both standard economic theory in the tradition of Adam Smith and the transaction costs theory have typically assumed that, with competitive supply conditions, market mechanisms overcome these coordination problems. The analysis offered above does not contradict that view. However, recent studies involving detailed models of market institutions for price and quantity determination do raise doubts about this assump-

Tracting relationships, the technological prowess of the supplier and the importance of the part supplied affect the bargaining power distribution among parties just as asset specificity does. If these factors are also similar to asset specificity in creating bargaining indeterminacy then they threaten to create coordination problems and may lead to the creation of specialized governance structures to alleviate the situation. In Japan, subcontractor associations (such as the Association of Toyota Suppliers) may fill this governance role.

tion, especially when more than two parties must agree in order to realize the benefits of exchange (Roberts, 1986). Of course, a key task of management is achieving coordination of actions within an organization, so recognizing the possibility of coordination failures' arising even in a competitive system of markets strengthens the case in favor of internal organization.

A second source of bargaining inefficiencies are measurement (information acquisition) costs. These costs have been emphasized by Barzel (1982) and by Kenney and Klein (1983) as a principal explanatory variable for specialized contracting practices and vertical integration. The idea is that individuals operating under standard short-term contracts will expend socially excessive amounts of resources on determining the private benefits and costs of an agreement when only its total costs and benefits (efficiency) matter.

As an example of how measurement costs affect market arrangements, both of the papers cited above discuss the Central Selling Organization (CSO) of the De Beers group, which in 1980 supplied 80-85% of the world market in diamonds. Kenney and Klein describe the CSO's marketing practices as follows:

"Each of the CSO's customers periodically informs the CSO of the kinds and quantities of diamonds it wishes to purchase. The CSO then assembles a single box (or "sight") of diamonds for the customer. Each box contains a number of folded, envelope-like packets called papers. The gems within each paper are similar and correspond to one of the CSO's classifications. The composition of any sight may differ slightly from that specified by the buyer because the supply of diamonds in each category is limited.

Once every five weeks, primarily at the CSO's offices in London, the diamond buyers are invited to inspect their

sights. Each box is marked with the buyer's name and a price. A single box may carry a price of up to several million pounds. Each buyer examines his sight before deciding whether to buy. Each buyer may spend as long as he wishes, examining his sight to see that each stone is graded correctly (that is, fits the description marked on each parcel). There is no negotiation over the price or composition of the sight. In rare cases where a buyer claims that a stone has been miscategorized by the CSO, and the sales staff agrees, the sight will be adjusted. If a buyer rejects the sight, he is offered no alternative box. Rejection is extremely rare, however, because buyers who reject the diamonds offered them are deleted from the list of invited customers.

Thus, stones (a) are sorted by De Beers into imperfectly homogeneous categories, (b) to be sold in preselected blocks, (c) to preselected buyers, (d) at non-negotiable prices, with (e) buyers' rejection of the sales offer leading to the withdrawal by De Beers of future invitations to purchase stones."

What accounts for these non-standard practices? If ordinary market arrangements were used, with the buyer and seller evaluating and haggling over each stone or group of stones, an inordinate amount of resources would be wasted in the evaluation and agreements might not always be reached. The buyer would carefully evaluate each rough stone to determine how to cut it to create the largest, most flawless and valuable cut diamond, and use that information to estimate the stone's value. To bargain effectively, the seller must be equally well informed, which would require a substantial non-productive investment on the seller's part. If the buyer and seller were to fail to agree on a price, another buyer would have to undertake the same evaluation, resulting in a duplication of effort and waste of resources.

³³Kenny and Klein (1983), page 502. Emphasis added.

Given De Beers's initial classification, there is little to be gained by further refinement in allocating diamonds among buyers. In a traditional market arrangement, customers would evaluate some stones that they will never cut and the seller, in self defense, would examine stones more closely than it would otherwise need to do. The De Beers system minimizes these measurement costs, which are attendant to haggling over price, and so represents one possible efficient response. The particular institution described also assigns most of the gains to De Beers, reflecting its monopoly position.

Notice how the De Beers system moves away from markets and introduces an element of centralization. Haggling is eliminated and the CSO is given authority to allocate the diamonds subject to certain categorization rules. If the buyer refuses his sight, he can terminate his relationship with De Beers. This is analogous to the right that an employee of any business organization has who is unhappy with his wage or job assignment; he can quit.

Even the most casual review of markets enables one to identify many circumstances in which pre-sale product evaluation and negotiation by the buyer would not help to allocate goods more efficiently but would give the buyer an edge in bargaining. In such circumstances, alternative arrangements that economize on these costs should be expected. Barzel uses this idea to explain the packaging of fruits and vegetables (which discourages product evaluation) and product warranties (which make careful product evaluations less

valuable to the buyer, and so reduce measurement activities).34 Kenney and Klein use it to explain the packaging of diamonds and the "block booking" of movies (which prevents theatre owners from picking and choosing among new releases and so economizes on measurement costs). The royalties paid to authors of books can be similarly explained. If fixed fees were paid to an author, the competing publishers would be led to incur excessive costs estimating the book's market potential for fear of the "Winner's Curse" according to which they acquire rights only to books of low potential that other publishers spurn.35 The use of royalties to compensate authors alleviates the Winner's Curse by making the publisher's payment depend on actual, rather than estimated, sales.36 Part of the costs of allowing speculators to trade in a commodity market is that their trading profits must compensate for their unproductive investments in the information that is so essential to them. 37 The fact that these last markets are auction markets with little explicit negotiations has little import for our argument.

may be inclined to economize on the newly unobservable quality. Warranties may provide a partial remedy for that, but they may be costly to exercise and may encourage customers to abuse goods.

³⁵Recall that our expansive definition of bargaining costs includes the costs of preparing bids in the event that there are several competing suppliers or buyers.

 $^{^{\}rm 32}$ The use of royalties also results in higher compensation, on average, for the author, even holding the publishers' information fixed. See Milgrom (1987).

⁹⁷See Hirschleifer (1971).

In general, measurable variety in the quality of an input increases the efforts that bargainers put into measurement activities, and hence increases the costs of market arrangements. Such diverse arrangements as vertical integration, product warranties, and non-standard market arrangements may emerge as the parties attempt to economize on these costs.

A third source of bargaining costs, and the one most often emphasized in the recent theoretical literature is private information about preferences. Parties may be delayed in reaching an agreement, or may even fail to reach an agreement at all, when one insists: "It's only worth \$50 to me and I won't pay a penny more," even though the actual value to the individual may in fact be far greater. Such misrepresentations are possible only to the extent that the seller is unsure of the buyer's actual valuation. Moreover, given uncertainty about whether trade is efficient, bargaining costs are inevitable, regardless of the bargaining procedure used. Intuitively, one expects that the greater the parties' uncertainty about one another's valuations of a proposed exchange, the greater the likely bargaining costs.

Our analysis of the sources of bargaining costs has been tentative and preliminary. Yet, it has served more than one

pages Sutton (1986) for a recent survey of non-cooperative barquining theory.

³⁹See Myerson and Satterthwaite (1983).

^{4°}To the best of our knowledge, however, this proposition has been nowhere formally established.

valuable purpose. It has reinforced the logic of transaction costs theory, provided a unifying perspective from which to investigate two previously distinct branches of transaction costs economics — that based on specialized assets and that based on measurement costs and pointed to a new agenda for bargaining theorists and experimenters.

4. Costs of Centralized Authority

Accounts of the economic growth of the West often emphasize the decentralization of economic control rights. As Rosenberg and Birdzell have recently written:

"We have emphasized the part played by innovation in Western growth. The decentralization of authority to make decisions about innovations, together with the resources to effectuate such decisions and to absorb the gains or losses resulting from them, merits similar emphasis as an explanation of Western innovation. This diffusion of authority was interwoven with the development of an essentially autonomous economic sector; with the widespread use of experiment to answer questions of technology, marketing, and organization for which answers could be found in no other way; and with the emergence of great diversity in the West's modes of organizing economic activity."⁴¹

Thus, Western economic history suggests that centralization stifles innovation. Is this a generalizable proposition? Even if one agrees that guild, church and feudal authorities squelched experimentation and innovation in Medieval Europe and that the mandarinate in China, the feudal lords in Japan, and the mullahs in the Islamic world had similar effects in their own domains, the historical evidence is still insufficient to establish that a modern

⁴¹Rosenberg and Birdzell (1986), page 24.

central planner, who has studied the lessons of history, cannot guide an economy to duplicate and improve upon the historical performance of market economies.

Why can't a centrally planned, consciously coordinated system always do at least as well as an unplanned, decentralized one? Some scholars, failing to find an answer to this question, have boldly (and we think, wrongly) concluded that there is no answer. In his 1928 Presidential Address to the American Economic Association, Frederick Taylor held that socialist economies can allocate goods as well as capitalist economies because they can duplicate those economies in all their desirable respects:

"In the case of a socialist state, the proper method of determining what commodities should be produced would be in outline substantially the same as that just described [for capitalist economies]. That is, the correct general procedure would be this: (1) The state would ensure to the citizen a given money income and (2) the state would authorize the citizen to spend that income as he chose in buying commodities produced by the state — a procedure which would virtually authorize the citizen to dictate just what commodities the economic authorities of the state should produce."

Substantially the same puzzle arises in trying to explain why there are any limits to the size and scope of firms. Thus, economists have asked: "Why, if by organising one can eliminate certain costs and in fact reduce the cost of production, are there any market transactions at all? Why is not all production carried out

⁴ªTaylor (1938), page 42.

by one big firm?"43 And, "Why can't a large firm do everything that a collection of small firms can do, and more?"44

We shall argue that there are predictable costs of centralized control - whether in firms or economic systems - that are not present in more decentralized organizations. Further, we shall take the point of view that what most distinguishes an actual firm from higher and lower level business organizations is the autonomy of its top manager, that is, the top manager's broad rights to intervene in lower level decisions and the relative immunity of his own decisions from intervention by others. From this perspective, the principles that guide the firm's decision to make or buy an input can be applied equally well to evaluate the productive efficiency of the capitalist and socialist economic systems.

Although our argument will treat the firm in a capitalist economy as a unit with considerable autonomy, one should not underestimate the degree of centralized authority present in market economies. Courts settle contract disputes and interpret the law. Government agencies issue permits, restrain certain business activities, and enforce orders of the courts. Legislatures enact

⁴³Coase (1937), page 395.

[&]quot;"Williamson (1986), page 131.

[&]quot;the distinguishing mark of the firm is the supersession of the price mechanism" and of Alchian and Demsetz (1972) that a firm is principally a "nexus of contracts."

⁴⁰A full comparison of the systems would of course entail much more, including their distributional, political, and moral characteristics.

taxes and laws to govern contracts, to limit firms' rights to pollute or engage in dangerous activities, to govern foreign trade, to control the use of land, and to promote societal ends such as development of the arts or improving the economic status of women and minorities. If our principles are indeed general ones, then these forms of centralized intervention must be subject to some of the same kinds of costs that accompany the creation of centralized executive authority in the firm.

Nor do we wish to overestimate the degree of centralized authority in firms. In its most decentralized, multidivisional form, business organizations allow division managers considerable autonomy from the central office. The holding companies that existed in the United States in the early twentieth century were even more decentralized; their central offices were little more than a substitute for the capital markets and bankers. However, in this essay, we shall regard these holding companies not as firms but instead as packages of separate, independent firms.

For our purposes, then, a firm is a business organization with a central office that retains substantial discretionary authority as well as substantial independence from other discretionary authorities. In capitalist economies, several institutions operate to support the much more extensive centralized control exercised by executives within firms than by courts or government agencies acting from outside the firm's boundaries. First, property rights tend to limit interventions by government agencies more than interventions by executives, because property rights generally reside at the

executive level or higher within the firm. Thus, a court, a regulatory agency, and a firm's central office can all order a polluting plant to cease operations until the problem is fixed, but the central office can also replace the plant manager with another manager if it finds that to be the most effective way to do the job. Second, executives generally have better and more fluid information systems than courts or government agencies do. Managers in firms hear most of the important information they need in conversations and meetings where they have the opportunity to query the source informally to resolve ambiguities and acquire needed detail 47: this stands in marked contrast to the written reports or adversary proceedings on which agencies and courts rely. Finally, executives can deliver incentives directly where they count most -- to individual employees -- and the incentives can take the form of rewards such as pay increases, bonuses, promotions, or desirable assignments, or the form of punishments (undesirable transfers or assignments or layoffs). The incentives offered by courts and government agencies consist mostly of threats to collect penalties against the firm's treasury.

Moreover, although some laws explicitly allow discretion to regulators, and others are so vague that the courts have considera-

⁴⁷Mintzberg (1973) summarizes the results of the many diary studies of how managers spend their time with the remark (page 36) that "gossip, speculation and hearsay form a most important part of the manager's information diet." Also (page 38), "Virtually every empirical study of managerial time allocation draws attention to the great proportion of time spent in verbal communications, with estimates ranging from 57 percent of time spent in face-to-face communication by foremen to 89 percent of episodes in verbal interaction by middle managers in a manufacturing company."

ble latitude to interpret them, the role of courts and government agencies is principally to enforce rules, and the court or agency must justify its action or decision in terms of the particular rule to be enforced. This procedure denies courts and agencies the kind of <u>fully discretionary</u> authority that the central office of a firm can exercise.

So, capitalist institutions support the structure of authority we have described as characteristic of the firm. Given this characterization, what are the diseconomies of organizing transactions internally within the firm? To bring the question into sharper focus, consider the case where a multidivisional conglomerate firm buys another firm and resolves to run it as a separate and independent division. This is a critical case for the theory, since such an acquisition represents a clear increase in centralization free of the confounding effects that come from the acquirer's attempts to integrate the acquired firm's assets and operations with its own. Experience shows that the performance of the acquired firm does sometimes deteriorate. By what mechanism does this deterioration occur?

The acquisition of Houston Oil and Minerals Corporation in late 1980 by Tenneco, Inc. (then America's largest conglomerate) is illustrative. Although Tenneco had resolved to run Houston as an independent subsidiary, before a year had passed, Tenneco "had lost 34% of Houston Oil's management, 25% of its explorationists, and 19%

[&]quot;The account given here is based on that given by Williamson (1986), page 158. Williamson cites an article in the Wall Street Journal of February 9, 1982 (page 17) as his original source.

of its production people, making it impossible to maintain it as a distinct unit." A Tenneco executive commented on the difficulties of bringing Houston, which was accustomed to giving large production-related bonuses to key people, into the conglomerate: "We have to ensure internal equity and apply the same standard of compensation to everyone." Why did this failure occur? And why did the executive insist on the need for "equity" and a commonly applied "standard of compensation"?

There are, we claim, three kinds of costs that quite generally accompany increases in discretionary centralized authority, and these apply in varying degrees to the case of conglomerate diversification. The first is that an individual whose authority has been increased may be unable to resist the temptation to interfere where he should not. This may happen simply because the individual feels an imperative to manage; that is what he is paid to do. Business people often cite this imperative to intervene as one cost of large government bureaucracies; the bureaucrats look for something to do whether their intervention is helpful or not. Or, the individual may have an actual personal interest in the decision: Will the empty lot next to the apartment building owned by the park commissioner's cousin be converted into a city park? Will the executive's protege be appointed to replace a retiring division head? For both of these reasons, one should expect that authority granted will be exercised more often and in other ways than efficiency considerations alone would dictate.

⁴⁹Quoted in the Wall Street Journal, February 9, 1982, page 17.

In the case of conglomerate integration, a previously separate firm, formerly protected from discretionary interventions by its property rights, becomes subordinated to a new executive authority. The individuals with newly augmented authority might seek to put their proteges into some of the high paying positions at the subsidiary, or they might seek to divert some of the subsidiary's capital budget to finance their own pet projects. On the other hand, a strong, honest central executive who is aware of this danger might sometimes thwart such attempts at interference.

There is a second kind of cost associated with discretionary authority that is incurred even when the central authority is both incorruptible and intelligent enough not to interfere in operations without good reason. A central office executive does not live on an island; he commonly relies extensively on others for information, suggestions, and analysis to reach effective decisions. If, as Milgrom (1986) holds, the employees affected by a decision are often the very ones on which the executive must rely to make the decision well, then the employees' efforts to influence the decision will impose costs on the organization. They may be led to distort the information they report, to withhold information from the central office and from other employees, and to spend valuable time polishing their credentials, establishing their qualifications for a desired assignment, or even just thinking about how they might

tions is to exploit the fact that many (virtually all) decisions require the participation of many individuals for their effectiveness."

influence some decision. Managers, worried about how higher authorities will evaluate their performance, may avoid risky investments that the firm should undertake because such investments pose career risks if they turn out badly. These distortions in the way employees spend their time, report their information, and make their decisions cause a loss of productivity, which is one category of what we call <u>influence costs</u>. They are a cost of discretionary authority because they arise only when there is an authority whose decisions can be influenced.

The costs of centralized authority so far described depend to some degree on the incentives that exist within the organization. Consequently, they can be at least partially alleviated by careful organizational design. Holmstrom and Ricart (1986) have investigated how a firm can, by carefully designed capital budgeting practices that reward investment and growth per se and establish high internal hurdle rates for investments, alleviate some of its managers' natural reluctance to undertake risky-but-profitable investments. Milgrom (1986) and Milgrom and Roberts (1987) have examined how compensation and promotion policies can be used to make employees more nearly indifferent about company decisions, so as to reduce resistance to change and other decision making distortions.

[&]quot;See Holmstrom (1982).

smlazear (1986) makes the related argument that wage "equality is desirable on efficiency grounds. The compression of wages suppresses unwanted uncooperative behavior" when employees compete for good jobs on the basis of comparative performance.

making. We have examined this option and its implications for corporate organization in detail in another essay, so we shall not consider it further here.

Japanese firms make some use of both wage policy and rules to facilitate extensive involvement of their employees in decision making without encouraging excessive influence activities. Their practices of lifetime employment for key decisionmakers, relatively small wage differentials within age cohorts and relatively low wages for senior executives, ³⁴ and promotions based in considerable degree on seniority ³⁵ combine to insulate employees from the effects of the firm's investment and promotion decisions and to make promotion decisions relatively immune to influence. Japanese firms, like their counterparts around the world, sometimes spin off or isolate unprofitable subsidiaries in order to prevent the unprofitable subsidiaries from imposing large influence costs on the organization

SaMilgrom and Roberts (1987).

compensation level at the top of the Japanese company is low — about \$100,000. This generally low level of pay for Japanese top management is borne out by other reports; for example Fortune of 19 March 1984 reported on 'salaries of Japanese chairmen and presidents, which range form \$50,000 to \$250,000 depending on company size.' Fortune also commented on U.S. executive salaries. 'In 1982 at least 85 American chief executives earned more than \$1 million.'" Abegglen and Stalk (1985), page 192.

is also a function of age in the Japanese firm, being provided within a predictable and narrow age range. Not everyone gets promoted—the escalator cannot carry everyone to the top floor—but promotion will rarely if ever take place until adequate seniority has been attained."

in attempts to claim corporate resources to cover their losses, and to force them to curtail operations. The Each of these devices to alleviate influence costs carries costs of its own. Leveling the wage distribution interferes with the other functions of wage policy; for example it increases the risk that valuable employees in key positions might quit. Emphasizing seniority over talent can result in a mismatch between employees and their jobs. Rewarding sheer investment can cause managers to spend too much time searching for more or bigger investments, rather than higher yielding ones. Isolating divisions narrows the range of options considered and may reduce valuable communications. All of these distortions and losses that firms suffer to reduce influence activities can themselves be categorized as influence costs.

In our example of the conglomerate merger, influence costs can affect both the new division head (the former chief executive of the acquired firm) and the old division heads. Either kind of division head may demand that the others purchase supplies from it. For example, a division head may argue that although his division's prices, based on average costs, make his product appear unattractive to the acquiring division, internal acquisition still serves the overall firm's interest because the marginal costs of production are low. Either kind of division head may play politics in an attempt

Japanese companies, Mitsubishi Chemical, Sumitomo Chemical, and Showa Denko, responded to the crisis in bauxite smelting caused by the oil price increases of 1973. "The aluminum producing units have been separated from their parent companies to isolate the problem and the losses, and their production facilities are being steadily reduced."

to influence job assignment, pay, and capital budgeting decisions. These are new and costly uses of at least the new division head's time that were not incurred in the same formor before the firm was acquired. Finally, division heads will expend some resources on defensive influence. For example, the newly acquired firm must be prepared to explain why its positions should be filled by promotion from within the division, or why its salaries and bonuses — high by comparison with those of employees in other divisions — should not be made part of the larger organization's general salary pool.

Taken together, the activities just described could consume a major portion of the time of division heads and central office personnel, diverting them from more productive activities. The boundaries that exist between independent firms serve to reduce the possibilities for influence. Consequently, those boundaries reduce influence costs.

The third category of costs associated with discretionary centralized authority are the costs of corruption (illicit influence costs): The central authority may seek bribes or other favors and may block efficient decisions when the bribes are not paid. Or, it may favor an inefficient supplier who has offered a bribe over a more efficient supplier who has not. Bribery scandals involving

⁼⁷⁰f course, the former chief executive may have spent some time with bankers or in attempts to impress his stockholders and directors. Our presumption is that he had greater autonomy from these authorities, and hence that he spent less time trying to influence them.

The prime example of this is selling. A good salesman is one who is successful at influencing the buyer's decision.

public officials are frequently reported in the press. Also reported are cases of sexual harassment where bosses demand sexual favors of candidates for promotion. Among the legal forms of bribery in the United States are the gifts given by many companies to the executives of their customers (unless the customer is a government entity). Wherever there is discretionary authority over decisions that people care about, there is a temptation to offer or solicit bribes.

In the Houston Oil and Mineral case described above, the most likely explanations of the failure of Tenneco to run Houston as an independent subsidiary lie in the first two categories given above. Tenneco's executives may have seen an opportunity to cut the wages or benefits of Houston's generously compensated professional workforce, disbelieving Houston's protestations that the results would be disastrous. Or, the employees of the other divisions may have coveted similar compensation packages, raising the costs to the organization of making an exception for Houston. Either way, the mere existence of an executive with discretionary authority to intervene imposed costs that would have been avoided if Houston had remained a separate firm.

So far, we have focused exclusively on the costs of discretionary centralized authority. There is another aspect of the centralization that is logically separate from authority to make discretionary interventions; it is the requirement that lower and middle level managers obtain authorization to undertake actions outside of some narrow, pre-approved set. Restrictions of this kind

may be an efficient way to provide incentives (Crawford and Sobel (1982)) or to improve coordination among diverse parts of the organization even when there is no incentive problem (Geanakoplos and Milgrom (1985)). However, to improve coordination by specifying that exceptions must be referred up the hierarchy (Galbraith (1977)) requires that the number of exceptions remain limited. The Since experiments and novel procedures are exceptional by nature and because they lack the urgency of exceptions arising in normal operations, the approval system limits the capacity of the organization to experiment and innovate.

There is one common exception to this rule; it arises when a manager is free to act without approval to make purchases or investments not exceeding some cash amount. If, when comparing lists of pre-approved actions within a given authority structure, we say that centralization is increased when the list is reduced, then we can understand a system of budget limits as a relatively decentralized version of the pre-approval system. Then, greater decentralization makes more experimentation and innovation possible.

Large modern firms often achieve a considerable degree of decentralization in this second sense. Units have great autonomy about how to spend their budgets and price their products. Some large firms, including IBM, DEC, and General Electric, have sometimes created special units to innovate new products and processes

The other apparent alternative -- expanding the central staff -- has costs of its own. It negates unity of control, leading to greater difficulties in assigning responsibilities for the performance of various units. It also makes it harder for top management to coordinate the handling of exceptions.

free from the normal system of approvals of the firm. Thus, although centralization in the sense of "approval systems" has generally been much higher within firms than in markets, we do not regard the costs of centralized approval systems as being inevitably incurred when a central discretionary authority is created.

5. Influence Costs in the Public Sector

Our theory of influence costs dovetails with recent economic theory of rent-seeking behavior. The seminal essays in this theory are those by Tullock (1967), Krueger (1974), and Posner (1975), all of which are reprinted in Buchanan, Tollison, and Tullock (1980). The theory holds that government interventions in the economy — whether in the form of tariffs, regulations, the awarding of monopoly franchises, or various attempts to correct market failures — are costly because they create rents and so lead firms and citizens to waste resources attempting to capture those rents. Its conclusion that rents lead to inefficiencies only when they result from government interventioner is, we believe, a mistake. Our general proposition is that any centralization of authority, whether in the public or private sector, creates the potential for intervention and so gives rise to costly influence activities, licit and illicit, and to excessive intervention by the central authority.

⁴⁰Buchanan (1980) advances this proposition.

These costs need to be weighed against the benefits of centralized authority to determine the efficient locus of authority.

Of course, our theoretical argument that increased centralization leads to increased influence applies with as much force to government and non-profit organizations as to firms. As an empirical matter, then, we should look for influence activities and their costs in the halls of government as well as in the executive offices of firms. Instances of influence in government are not difficult to find. The frustration of U.S. federal government officials who try to manage the nation's affairs in the face of constant attempts at influence was highlighted in recent testimony by the U.S. Secretary of State:

"Nothing ever gets settled in this town. It's not like running a company or even a university. It's a seething debating society in which the debate never stops, in which people never give up, including me, and that's the atmosphere in which you administer."

The current crisis in tort litigation in the United States provides a second illustration of the importance of influence costs in government. The crisis has arisen in part from the increasing frequency with which novel legal arguments win. In effect, courts

reduce or eliminate bargaining costs. Also, as Milgrom (1986) and Milgrom and Roberts (1987) have argued, the influence activities themselves may have a role to play in improving decision making, since the influencing parties may have information and suggestions that will be of value. In similar fashion, lobbyists and other advocates may contribute to better government decisionmaking, and that possibility must be weighed in any fair-minded evaluation of the economics of government intervention.

George Schultz, U.S. Secretary of State, testifying to the House Foreign Affairs Committee, December 8, 1986, as reported in the Oakland (California) Tribune, December 9, 1986.

are increasingly acting like discretionary authorities, and litigants are incurring costs in their efforts to capture the newly
appropriable sums. The costs to the United States of this litigation, in which some of the nation's finest minds are diverted into
largely non-productive legal activities and talented corporate
executives are led to devote a large part of their time to defending
and avoiding lawsuits, are enormous. The offsetting gains, for
example in terms of improved justice, are much harder to estimate.
Limits on damage awards, which are puzzling in standard economic
theory, are easily understood as a device to reduce influence
costs.

The importance of limits on the discretionary authority of government for encouraging economic development is clear in the economic history of Western Europe. Rosenberg and Birdzell have identified these limits as among the key factors that encouraged the development of trade and early capitalism:

"Some of the institutional innovations reduced the risks of trade, either political or commercial. Among them were a legal system designed to give predictable, rather than discretionary, decisions; the introduction of bills of exchange, which facilitated the transfer of money and provided the credit need for commercial transactions; the rise of an insurance market; and the change of governmental revenue systems from discretionary expropriation to systematic taxation — a change closely linked to the development of the institution of private property." 64

Description that the court will sometimes fail to find fault where it should and since damage awards must provide the economic incentive to take care to protect against losses, the award should be larger than the actual damage suffered when the party is found guilty. It is this logic that underlies the treble damage rule in antitrust.

^{**}Rosenberg and Birdzell (1986), page 113.

6. Summary and Conclusions

We have examined the organization of economic activity under the hypothesis that capitalist economic institutions are so organized as to minimize the sum of the costs of resources used in production and the costs of managing the necessary transactions.

The costs of negotiating short term contracts emerged as the distinctive costs of the traditional market mode of transaction. An analysis of the determinants of these bargaining costs indicated that two leading theories — one attributing the market transaction costs primarily to specialized assets and the other attributing them primarily to costs of measurement — could both be subsumed under the bargaining cost approach.

The costs associated with non-market forms of organization have received little attention in the existing literature, but must be assessed to identify when the market form of organization is more economical than internal procurement. As a first step, we argued that transactions within firms in a capitalist economy are characterized by greater centralization of authority than market mediated transactions. Indeed, the autonomy and discretion of top management and the lesser autonomy of lower levels of management is the principal defining characteristic of the firm.

Whenever a central authority with discretion to intervene exists, whether that authority is a governmental unit or an executive within a firm, certain identifiable costs are incurred. These include (1) a tendency for excessive intervention by the authority,

both because intervening is his job and because he may have a personal interest in certain decisions, (2) an increase in the time devoted to influence activities, as interested parties seek to have the authority intervene in particular ways or adopt their favored alternatives, and (3) a degradation of decision making resulting from bribery and other illicit influence activities.

Another important kind of centralization in organizations occurs when lower and middle level managers must have approval from one of a small number of executives to employ any exceptional procedures. Centralization of that kind inhibits and delays experimentation and innovation. It delays by adding another step to the process of implementing an innovation or experiment. It inhibits by creating a bottleneck that limits the rate at which non-standard procedures can be introduced. However, unlike the other consequences of creating a central discretionary authority, this one can be and sometimes has been avoided by the simple expedient of giving broad budget authority to department or division managers while still retaining the right to intervene as needed.

References

- Abegglen, James C. and George Stalk, Jr., <u>Kaisha: The Japanese</u>
 <u>Corporation</u>, New York: Basic Books, 1985.
- Alchian, Armen and Harold Demsetz, "Production, Information Costs, and Economic Organization," <u>American Economic Review</u> 62, 1972: 777-795.
- Arrow, Kenneth, The Limits of Organization, New York: Norton, 1974.
- Asanuma, Banri, "Traditional Structure of Parts Supply in Japanese Automobile and Electric Machinery Industries: A Comparative Analysis," Technical Report No. 1, Socio-Economic Systems Research Project, Kyoto University, July 1985.
- Barzel, Yoram, "Measurement Cost and the Organization of Markets,",

 Journal of Law and Economics 25, 1982:27-48.
- Buchanan, James, "Rent Seeking and Profit Seeking" in <u>Toward a</u>

 <u>Theory of the Rent Seeking Society</u>, College Station, Texas:
 Texas A&M University Press, 1980.
- of the Rent Seeking Society, College Station, Texas: Texas A&M University Press, 1980.
- Chandler, Alfred D., Jr., <u>Strategy and Structure</u>, Cambridge, Mass.: MIT Press, 1962.
- tion in American Business, Cambridge, Mass.: The Belknap Press, 1977.
- Coase, Ronald, "The Nature of the Firm," Economica, 1937:386-405.
- ------, "The Problem of Social Cost," <u>Journal of Law and</u>
 <u>Economics</u>, 1960:1-44.
- Crawford, Vincent, "Long Term Relationships Governed by Short Term Contracts," Working Paper No. 205, Industrial Relations Section, Princeton University, 1986.
- Debreu, Gerard, <u>Theory of Value</u>, New Haven: Yale University Press, 1959.

- Farrell, Joseph and Nancy Gallini, "Second-Sourcing as a Commitment: Monopoly Incentives to Attract Competition," University of California Department of Economics Working Paper No. 8618, 1986.
- Fudenberg, Drew, Bengt Holmstrom and Paul Milgrom, "Short Term Contracts in Long Term Relationships," mimeo, 1987.
- Galbraith, Jay, <u>Organization Design</u>, Reading, Mass: Addison-Wesley, 1977.
- Geanakoplos, John and Paul Milgrom, "A Theory of Hierarchies Based on Limited Managerial Attention," Cowles Foundation Discussion Paper No. 775, 1985.
- Grossman, Sanford and Oliver Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,"

 Journal of Political Economy, 1986: 691-719.
- Grout, Paul, "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," <u>Econometrica</u> 52, 1984:449-460.
- Hirschleifer, Jack, "The Private and Social Value of Information and the Reward to Inventive Activity," American Economic Review 61, 1971:561-574.
- Holmstrom, Bengt, "Managerial Incentives -- A Dynamic Perspective," in <u>Essays in Economics and Management in Honor of Lars Wahl-beck</u>, Helsinki, Finland: Swedish School of Economics, 1982.
- Management," Quarterly Journal of Economics 101, 1986:835-860.
- Kenney, Roy and Benjamin Klein, "The Economics of Block Booking,"

 <u>Journal of Law and Economics</u> 26, 1983:497-540.
- Klein, Benjamin, Robert Crawford and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," <u>Journal of Law and Economics</u> 21, 1978:297-326.
- Krueger, Anne, "The Political Economy of the Rent-Seeking Society,"

 American Economic Review 64, 1974:291-303.
- Lazear, Edward, "Pay Inequality and Industrial Politics," Hoover Institution Working Paper No. E-86-12.
- Lindblom, Charles E., "The Science of 'Muddling Through'", <u>Public Administration Review</u> 19, 1959:79-88.

- Milgrom, Paul, "Auction Theory," <u>Advances in Economic Theory, 1985:</u>

 <u>Invited papers for the Fifth World Congress of the Econometric Society</u>, Truman Bewley (ed), London: Cambridge University

 Press, 1987.
- -----, "Quasirents, Influence, and Grganizational Form,"

 Yale School of Organization and Management Working Paper, 1986.
- ----- and John Roberts, "An Economic Approach to Influence Activities and Organizational Responses," unpublished working paper, 1987.
- Mintzberg, Henry, The Nature of Managerial Work, New York: Harper and Row, 1973.
- Myerson, Roger and Mark Satterthwaite, "Efficient Mechanisms for Bilateral Trading," <u>Journal of Economic Theory</u> 28, 1983:265-281.
- Nash, John, "The Bargaining Problem," Econometrica 18, 1950:155-162.
- -----, "Two Person Cooperative Games," <u>Econometrica</u> 21, 1953:128-140.
- Nelson, Richard and Sidney Winter, <u>An Evolutionary Theory of</u>
 <u>Economic Change</u>, Cambridge, Mass.: Harvard University Press,
 1982.
- Posner, Richard, "The Social Costs of Monopoly and Regulation,"

 <u>Journal of Political Economy</u> 83, 1975:807-27.
- Quine, Willard, Word and Object, New York: Wiley, 1960.
- Roberts, John, "An Equilibrium Model with Keynesian Unemployment at Walrasian Prices," Stanford GSB Research Paper No. 908, 1986.
- Rosenberg, Nathan and L.E. Birdzell, How the West Grew Rich, New York: Basic Books, 1986.
- Roth, Alvin E., "Bargaining Phenomena and Bargaining Theory," unpublished working paper, 1986.
- Schelling, Thomas, <u>The Strategy of Conflict</u>, Cambridge, Mass: Harvard University Press, 1960.
- Shepard, Andrea, "Licensing to Enhance Demand for New Technologies," unpublished working paper, Yale University, 1986.
- Sutton, John, "Noncooperative Bargaining Theory: An Introduction,"

 Review of Economic Studies 53, 1986:709-724.

- Taylor, Fred M., "The Guidance of Production in a Socialist State,"

 <u>American Economic Review</u> 19, 1929. Reprinted in <u>On the</u>

 <u>Economic Theory of Socialism</u>, New York: McGraw Hill, 1938.
- Tirole, Jean, "Procurement and Renegotiation," <u>Journal of Political</u> Economy 94, 1986:235-259.
- Tullock, Gordon, "The Welfare Costs of Tariffs, Monopolies, and Theft," Western Economic Journal 5, 1967:224-32.
- Williamson, Oliver, The Economic Institutions of Capitalism, New York: The Free Press, 1985.

RECENT ISSUES OF THE WORKING PAPER SERIES OF THE DEPARTMENT OF ECONOMICS UNIVERSITY OF CALIFORNIA, BERKELEY

Copies may be obtained from the Institute of Business and Economic Research. See the inside cover for further details.

- 8618 Joseph Farrell and Nancy T. Gallini SECOND-SOURCING AS A COMMITMENT: MONOPOLY INCENTIVES TO ATTRACT COMPETITION Dec-86.
- 8619 Roger L. Ransom and Richard Sutch
 TONTINE INSURANCE AND THE ARMSTRONG INVESTIGATION: A CASE OF STIFLED INNOVATION
 IN THE AMERICAN LIFE INSURANCE INDUSTRY, 1868-1905
 Dec-86.
- 8620 Joseph Farrell and Robert Gibbons CHEAP TALK IN BARGAINING GAMES
 Dec-86.
- 8621 Robert M. Anderson
 CORE ALLOCATIONS AND SMALL INCOME TRANSFERS
 Dec-86.
- 8722 Joseph Farrell COMPETITION WITH LOCK-IN Jan-87.
- 9723 Pranab Bardhan and Nirvikar Singh MULTINATIONAL RIVALRY AND NATIONAL ADVANTAGE: SOME THEORETICAL CONSIDERATIONS

 Jan-87.
- 8724 Albert Fishlow
 LESSONS OF THE 1890S FOR THE 1980S
 Jan-87.

RECENT ISSUES OF THE WORKING PAPER SERIES OF THE DEPARTMENT OF ECONOMICS UNIVERSITY OF CALIFORNIA, BERKELEY

Copies may be obtained from the Institute of Business and Economic Research. See the inside cover for further details.

- 8725 Jeffrey A. Frankel
 THE SOURCES OF DISAGREEMENT AMONG INTERNATIONAL MACRO MODELS,
 AND IMPLICATIONS FOR POLICY COORDINATION
 Jan-87.
- 8726 Jeffrey A. Frankel and James H. Stock REGRESSION VS. VOLATILITY TESTS OF THE EFFICIENCY OF FOREIGN EXCHANGE MARKETS Jan-87.
- 8727 Joseph Farrell and Carl Shapiro DYNAMIC COMPETITION WITH LOCK-IN Jan-87.
- 8728 James A. Mirrlees ECONOMIC POLICY AND NONRATIONAL BEHAVIOUR Feb-87.
- 8729 Jeffrey A. Frankel
 THE IMPACT OF U.S. ECONOMIC POLICIES ON A COMMODITY-EXPORTING DEBTOR:
 THE CASE OF THAILAND
 Feb-87.
- 8730 Richard J. Gilbert and Marvin Lieberman INVESTMENT AND COORDINATION IN OLIGOPOLISTIC INDUSTRIES Feb-87.
- 8731 Paul Milgrom and John Roberts
 BARGAINING AND INFLUENCE COSTS AND THE ORGANIZATION OF ECONOMIC ACTIVITY
 Feb-87.