Explaining the Swollen Middle: Why Most Transactions Are a Mix of "Market" and "Hierarchy"

Author(s): Jean-Francois Hennart


Published by: INFORMS

Stable URL: http://www.jstor.org/stable/2635079


Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode= informs.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to Organization Science.
EXPLAINING THE SWOLLEN MIDDLE: WHY MOST TRANSACTIONS ARE A MIX OF "MARKET" AND "HIERARCHY"*

JEAN-FRANCOIS HENNART

University of Illinois at Urbana-Champaign, Department of Business Administration, 350 Commerce West, 1206 South Sixth Street, Champaign, Illinois 61820

Why are firms sometimes more efficient than markets at organizing transactions? Why are most transactions arrayed neither at the pure "market" nor at the pure "hierarchy" end of the continuum, but rather in the "swollen middle," incorporating features of both "market" and "hierarchy"? Why don't firms make greater use of price incentives? This paper addresses these three questions by developing a model of the choice of institution.

One key building block is the distinction between organizing methods (hierarchy and the price system) and institutions (firms and markets). Hierarchy and the price system are two distinct methods for organizing transactions, each with particular costs and benefits. Markets and firms are institutions which use one or both of these methods. Although markets predominantly use prices and firms rely principally on hierarchy, there is not a one-to-one correspondence between prices and markets or between hierarchy and firms. Indeed, the paper argues that it is generally more efficient to use a mix of both methods than to specialize in either.

The paper focuses on the enforcement properties of prices and hierarchy. Hierarchy controls individuals directly by constraining their behavior (by imposing behavior constraints) while prices do it indirectly by measuring their outputs (through price constraints). Under hierarchy, individuals receive a salary to do as told, while self-employed individuals governed by the price system are rewarded on the basis of their output. Each system has its own biases: using prices maximizes effort (minimizes shirking) but incites individuals to inflate the price and/or reduce the quality of their output. (It encourages cheating.) Relying on hierarchy results in the opposite bias: under hierarchy individuals are not paid in function of their output, but instead are rewarded for following directives. They have, thus, strong incentives to minimize effort (to shirk) unless properly supervised, but, being paid a fixed sum to follow orders, they have few incentives to cheat. Hence the price system experiences low shirking, but potentially high cheating costs, while hierarchy faces low cheating but high shirking costs.

Organizing costs are the sum of shirking and cheating costs. Any given transaction will be organized by the mix of price and hierarchy (i.e., by the mix of price and behavior constraints) that minimizes organizing costs. A transaction will be organized within a firm if the reduction in cheating costs achieved by replacing price constraints by behavior constraints exceeds the resulting increase in shirking costs and by the market in the opposite case.

The paper shows that cheating and shirking costs increase more than proportionately as one concentrates in either pure price or behavior constraints. Hence using a mix of both methods generally minimizes the sum of cheating and shirking costs. This explains why most transactions exhibit features of both markets and hierarchy. The paper shows clearly the tradeoff involved between price and behavior constraints. It explains the costs and benefits of using two types of price incentives in firms, piecework and profit centers, and predicts when they will be used.

(Accepted by Arie Y. Lewin; received February 27, 1991. This paper has been with the author for two revisions.)
institutions, both domestic and international. Yet the paradigm is not without its critics. This paper addresses two fundamental criticisms leveled against transaction costs theory. First, transaction cost theory is only a theory of market failure—it does not explain why firms succeed (Demsetz 1988). Second, as Perrow (1986) asserts, the categories of “market” and “hierarchy” used in transaction costs theory are not useful; observation shows that most transactions cannot be categorized as either “pure market” or “pure hierarchy” (Powell 1987; Stinchcombe 1990). This paper extends transaction cost theory to show why “pure hierarchy” can sometimes be a more efficient organizing method than “pure market” and how this explains why these two pure types often comingle in real life.

The tendency of some of the writings in TCE has been to rely on the presence of transaction costs in markets to explain why firms internalize a transaction. It seems clear, however, that the presence of market transaction costs is not a sufficient condition for internalization. Organizing exchange and cooperation is costly, and while TCE has mostly focused on the costs of organizing transactions in markets (market transaction costs), firms also incur organizing costs. It is, in fact, perfectly plausible that firms, in their attempt to internalize market transactions, experience higher organizing costs than markets. If organizing costs in firms are so high that they absorb all of the gains from exchange and coordination, then no economic interaction will take place, either within firms or in markets. A complete theory of economic institutions should therefore consider simultaneously the costs of organizing transactions in markets (transaction costs) and those of effecting exchange within the firm (management costs) and should explain how firms can achieve lower organizing costs than markets. This requires a theory of the organizing method(s) used by firms, and of the costs and benefits of those methods relative to those used in markets.

Demsetz (1988, p. 149) has argued that TCE has not provided such a theory because it does not make a clear distinction between transaction cost and management costs, and that this failure to distinguish between the two has limited its applicability to substantive issues:

It is hard to know whether we are dealing with a transaction or a management cost until we already know whether we are discussing a firm or a market. . . . This makes it difficult to use the magnitude of “transaction” relative to “management” cost to predict how changed circumstances affect economic organization.

Another major criticism of TCE is that it neglects the complexity of actual institutions by focusing on the two extremes of markets and hierarchy. According to Perrow (1986, p. 255),

There are strong elements of markets within hierarchies. On the other hand, markets have strong elements of hierarchy within them. The distinction between markets and hierarchies is greatly overdrawn. The continuum from market to hierarchy is less like a ruler than a football, with a vanishing small pure type at each end, and a swollen middle that mixes the two (as Sidney Winter once put it in a seminar). . . . Very little of organizational life remains at the two ends of the ‘football’.

Demsetz’s and Perrow’s criticisms are linked. If TCE cannot conceptually distinguish between the costs of using the price system and those of using hierarchy, then it will have difficulty in explaining why both organizing methods are often used simultaneously.

The following pages show how transaction cost theory can be extended to respond to these critiques. Section 2 outlines the differences in the two polar organizing methods of price and hierarchy. Armed with a better understanding of how the costs
of using the price system differ from those of using hierarchy we can tackle the issue raised by Perrow. Section 3 shows that the hybrid arrangements that make up the “swollen middle” can be analyzed as a tradeoff between these two types of cost. Section 4 discusses the limits to the use of price incentives in firms. Section 5 summarizes the argument.

The comparative institutional model presented here builds upon transaction costs and organizational economics (Williamson 1975, 1985; Ouchi 1977, 1979; McManus 1972; Hennart 1982, 1986; Barney and Ouchi 1986; Masten 1988), property rights theory (Alchian and Demsetz 1972; Alchian 1984; Grossman and Hart 1988; Demsetz 1988; Barzel 1989) and agency theory (for surveys see Eisenhardt 1985, 1989; Levinthal 1988), but differs from these approaches in some fundamental ways. In contrast to some agency models—for example Eisenhardt (1985)—it does not assume risk aversion on the part of agents and principals. In contrast to Ouchi (1979) and Eisenhardt (1985), organizing costs are more than just the costs of measuring outcomes or behaviors; instead they are the sum of measurement costs and of the costs that result from imperfect measurement. While “shirking” has generally been considered as one of those costs (Alchian and Demsetz 1972), the model highlights the symmetrical category of “cheating.”

In common with most of the literature cited above, the model relies on a number of assumptions which have been criticized in the sociological literature (Granovetter 1985; Etzioni 1988; Mansbridge 1990). Although some of these criticisms are discussed in the following pages, no attempt is made to systematically incorporate them into the model.

2. A Model of the Choice between Firms and Markets

This section develops a model of the factors influencing the choice between firms and markets based on a comparison of the costs of cheating and shirking. The argument consists of six basic propositions:

(1) One must distinguish between methods of organizing (the price system and hierarchy) and economic institutions (markets and firms). There is no one-to-one correspondence between the two, and any given institution may, under specific circumstances, use a mix of both methods of organization.

(2) The two organizing methods, the price system and hierarchy, use different techniques to organize economic activities. The price system rewards agents on the basis of their outputs; hierarchy rewards on the basis of behavior (inputs). In a world of zero transaction costs, both would be equally effective (Coase 1937). With positive organizing costs, each technique will experience divergent levels of organizing costs for a given transaction.

(3) The cost of using price constraints (cheating cost) is the cost of measuring output, plus the losses due to fraud when measurement is imperfect. The cost of using hierarchy is that of using behavior constraints. This cost, which we call “shirking cost,” is the sum of the cost of constraining behavior plus the residual amount of shirking due to imperfect behavior constraints.

(4) Price constraints minimize shirking but encourage cheating; behavior constraints minimize cheating but encourage shirking. The choice between using prices and hierarchy will depend on the relative costs of measuring output plus that of tolerating the residual amount of cheating as compared to those of constraining behavior and of bearing the residual amount of shirking.

(5) Markets are institutions that predominantly use the price method of organizing. Firms predominantly rely on hierarchy. However, because of diminishing returns to measuring output and constraining behavior, both firms and markets will often use a
mix of price and behavior constraints. The firm's mix will contain a high proportion of behavior relative to price constraints; the mix in markets will be biased towards price constraints.

(6) The combination of price and behavior constraints defines a wide variety of institutional forms along a continuum which goes from pure spot markets to traditional firms. The model explains why the most common institutional forms use both methods of organizing.

These propositions are more fully developed in the pages that follow.

2.1. Organizing Methods vs. Economic Institutions

The starting point of our analysis is the distinction between methods of organizing (the price system and hierarchy) and economic institutions (markets and firms). Prices and what we call "hierarchy" are methods used to organize economic activities. Markets and firms are institutions which use one or both methods to achieve that goal. Although markets predominantly use prices as methods of organizing and firms rely principally on hierarchy, there is not a one-to-one correspondence between prices and markets and hierarchy and firms. Thus firms may develop internal pricing systems and markets may use behavioral constraints. Indeed, as we will show below, it is generally efficient to use a mix of both methods of organizing in firms as well as in markets.

2.2. Prices and Hierarchy in the Absence of Organizing Costs

Cooperation between individuals can be productive, either because some tasks are best achieved through pooling of effort or because individuals have differing abilities and they can exploit these differences through exchange. Three tasks are required to achieve cooperation. Parties to the interaction must be made aware of the potential gains of cooperation. The gains of cooperation are joint, so a formula for dividing those gains must be imposed on the parties to curb bargaining. This sharing rule must be enforced.

Achieving cooperation would be costless if individuals had unlimited abilities, the same goals, and were perfectly selfless. We assume that individuals have "bounded rationality," that they have only partially overlapping goals, and that they are opportunistic (Williamson 1975). These assumptions are crucial to the model, for without them, organizing costs would be zero.

Williamson (1985, p. 47) has defined opportunism as "self-interest seeking with guile." Williamson's assumption of opportunism does not require that everyone be opportunistic or that opportunism be a constant. Devising complex institutions to protect against opportunism is costly; it is therefore advantageous to signal honesty, and in the long-term interest of the parties to behave honestly when such behavior is likely to provide a credible signal (Spence 1974, Elster 1990).1 Hence an important check on opportunism is its impact on reputation. Since the structure and the stability of the social network in which transactions are embedded affect the payoff from reputation, they also affect the level of opportunism. The social network also generates powerful pressures that limit opportunism, a point which has been stressed in the sociological literature (e.g., Granovetter 1985, Jencks 1990, Etzioni 1988) and acknowledged by Williamson (1985, p. 22).2 Hence we would expect opportunism to

---

1 Individuals will behave honestly if it is costly to appear honest while being dishonest (Ackerlof 1983).
2 Jencks (1990, p. 54) identifies three factors that limit opportunism: empathy, community, and morality. "Empathy" involves identification with specific individuals, "communitarian unselfishness" involves identity with a collectivity, "morality" involves the incorporation of external moral ideals derived from the collective culture of the larger group into one's sense of self.
vary across individuals, transactions, industry conditions, and countries. However, if it is costly for human decision makers to distinguish *ex ante* between those who will behave opportunistically and those who will not, "prudence suggests the adoption of the assumption of opportunism" (Barney 1990) and the design of institutions will often reflect that fact (Williamson 1974, p. 27).

There are two basic ways of reconciling the interests of opportunistic individuals (Ouchi 1979). The first one, which can be called "internal control," consists in selecting individuals who have the same goals or in investing resources to persuade individuals who have different goals to internalize one particular set of goals, so that the goals of all members are now congruent. This has been called the "clan" solution (Ouchi 1979). The second solution is to use external control mechanisms. Hierarchy and the price system are two such mechanisms. Internal and external control are substitutes and conceptually distinct, although all institutions use a mix of these two modes of control (Ouchi 1979). This paper will not discuss the choice between internal and external control, but will focus instead on the choice between two modes of external control, the price system and hierarchy.

If we assume that individuals are opportunistic, boundedly rational, and have different goals, informing parties, rewarding them for productive behavior, and curbing bargaining will be costly. These information, bargaining, and enforcement costs are the costs of organizing economic activities. They must be borne when using either method of organizing, whether it is the price system or hierarchy. Because the price system and hierarchy use different incentive and information structures, they experience divergent costs when organizing a given transaction. Each method will therefore have a comparative advantage in organizing a particular set of transactions (Ouchi 1979). Assuming some degree of competition which penalizes the use of inefficient methods, the method that will be used to organize a transaction will be the least costly (Williamson 1975).

The following paragraphs briefly describe how the price system and hierarchy organize economic activities in the absence of bounded rationality and opportunism. This allows us to separate the description of the basic method of organizing from the actual performance of these methods of organizing when organizing costs are positive. To simplify our discussion, the emphasis will be on enforcement costs, although information and bargaining costs will also be discussed. Let us first turn to the price system.

Prices perform the three tasks of organizing: they inform parties, they curb bargaining, and they provide rewards and punishments. In the absence of transaction costs (when there are perfectly defined property rights with costless enforcement,
zero information costs, and a large number of buyers and sellers), prices convey information on the consequences of one’s actions so as to allow parties to reach optimal decisions on the allocation of tasks.\(^7\) The information structure of a market is fully decentralized, and each party receives information through prices about everyone else’s needs and desires, and adapts to it in a way that maximizes joint welfare. When markets function in this way, prices are exogenous and bargaining doesn’t pay. Prices meter and perfectly reward an agent’s behavior; the gross rewards that agents receive are a direct function of their output times those market prices (Arrow 1974, Ouchi 1979).

In the absence of organizing costs, hierarchy would also organize economic activities perfectly. The hierarchical method of organizing is characterized by centralized information and the use of behavior constraints. (Recall that we define hierarchy as a method of organizing—hierarchy is not synonymous here with “firm” nor with “upper level managers”\(^7\).) Thus, while information is decentralized with prices, it is centralized with hierarchy. Under hierarchy individuals (employees) are asked to channel the information they possess to a central party who assimilates this information and retransmits relevant parts of it back to the employee in the form of directives. With unbounded rationality, this is as efficient a method of making optimal decisions on the allocation of tasks as the decentralized system of market prices (Williamson 1975).

Individuals who organize their interactions through the price system collect their own information and make their own productive decisions. They are rewarded in proportion to their output measured at market prices. Their rewards serve as incentives to collect the best information possible and to act on it. By contrast, under hierarchy, the individual agrees to have the central party (the boss) direct the allocation of his or her own resources (such as his or her labor-time and effort). The employee agrees to do as told, within the constraints of social custom. Under these conditions, no rational individual would permit someone else to direct his productive behavior if he was rewarded by his output measured at market prices. He then would have to bear the cost of being ordered to perform tasks that do not maximize his income. Consequently, employees are not rewarded by their output measured at market prices, but instead by their obedience to managerial directives. In other words, while prices indirectly guide behavior by rewarding output, hierarchy directly controls individuals by specifying behavior and rewarding compliance. Because employees are not rewarded by their output, they will be less concerned about the allocation of their resources, and they will accept management’s right to direct their activities by fiat.

2.3. Prices and Hierarchy with Positive Organizing Costs

With bounded rationality, opportunism, and a divergence of goals between individuals, the price system and hierarchy will both incur costs in organizing economic activities. But because of the differences in the way they perform this task, each method of organizing will incur a different type of costs. The efficiency of the price system will be reduced by the cost of cheating, while hierarchy will suffer shirking costs.

For prices to guide individual action efficiently to reap the benefits of exchange and coordination, they must reflect the value of goods and services. This in turn presup-

---

\(^7\)This assumes that preferences are given, an assumption criticized by Galbraith (1967). Although it is crucial to the Arrow-Debreu theorem (Arrow and Debreu 1954), it is not clear that it is required to say that the price system can be efficient in a comparative-institutional sense (i.e., relative to other methods of organizing).
poses that the value of the output produced can be perfectly measured by the interacting parties. In reality, bounded rationality will make measurement costs positive, and output will be imperfectly measured. Hence prices will not provide the “right” signals; agents will be incited to use too much of the goods whose price is below cost, and too little of the goods priced below their benefit to the group. Similarly, when the number of buyers and sellers falls, prices are no longer exogenous, and it will pay for agents to invest in bargaining. When the number of buyers and sellers is large and consumers are perfectly informed, competition will weed out dishonest traders, since they will not get repeat business. A small number of parties (a thin market), on the other hand, makes it difficult to switch partners, while bounded rationality clouds the link between performance and reputation. Both of these factors will reduce the self-enforcing properties of markets.

Because measuring outputs experiences diminishing returns, it will not pay to attempt to measure outputs perfectly. Measurement costs will be incurred up to the point where the marginal cost of measurement equals its marginal benefit. This will make it possible for traders to cheat. For example, farmers may surreptitiously add water to the milk they sell to dairies, or may sell milk contaminated with pesticides, and yet charge dairies the price of unadulterated milk. The dairy could perform extensive tests on each batch of milk, but the cost is likely to be prohibitive. Instead, it may perform limited tests on some of the milk it purchases, and accept some cheating because it is not worth curbing. The organizing costs incurred by the price system (as far as enforcement is concerned) are referred to as cheating costs, and they are the sum of the cost of measuring output plus the cost of the residual amount of cheating due to imperfect measurement.

When cheating costs are high, a switch to hierarchy may reduce organizing costs. Hierarchy reduces cheating by reducing the incentives individuals have to cheat. It does this by breaking the connection between outputs and rewards. Individuals are paid a fixed sum to follow orders, and thus gain little from cheating.

Such a system of rewards independent from outputs has one unavoidable consequence: while it reduces cheating, it also reduces incentives to work. When agents governed by prices (self-employed individuals) take the day off, their output suffers. Since lowered output decreases their reward, they pay the full cost of their loafing. Employees, on the other hand, have less incentive to generate output, because their reward is no longer directly proportional to that output measured at market prices. As long as their behavior is costly to monitor, employees will have incentives to shirk, i.e., to break the spirit or the letter of their promise to do as told. Note that shirking does not necessarily mean loafing. It means that the behavior of an employee will differ from what it would be if he were self-employed. Employers will therefore have to invest resources to monitor and direct the behavior of employees. The amount of shirking will depend on the extent to which the employee’s goals differ from those of the employer and on the cost of constraining employee behavior (Eisenhardt 1985).

8In some cases measurement costs may be so high that some goods cannot be priced. They will carry a zero price, as with clean air or noise. This constitutes an extreme case of “market failure.”

9Shirking takes place when the employee offers “perfunctory” as opposed to “consummate” cooperation. Perfunctory cooperation is compliance with directives, while consummate cooperation involves accepting responsibilities and exercising initiative (Blau and Scott 1962, p. 140; Williamson 1975, p. 69). The problem of shirking becomes massive in the case of slave labor; it is very difficult to have slave labor perform tasks that require care and initiative (Genovese 1967, Fenoaltea 1984).

10Reputation will constrain shirking by employees. There are three main reasons why reputation fails to eliminate shirking. First, agents are only boundedly rational; they are imperfectly informed, and they forget. Second, observed performance is the joint outcome of effort and luck, and it is not always easy to separate the two. Lastly, some tasks require joint effort, making it difficult to isolate individual contributions (Alchian and Demsetz 1972).
There are likely to be diminishing returns to constraining employee behavior, so it will not pay to totally eliminate shirking. Shirking costs, one of the costs of using hierarchy, will therefore be the sum of the costs of constraining behavior and of those of bearing the residual amount of shirking. In summary, the two methods of organizing, the price system and hierarchy, are essentially substitutes: switching from prices to hierarchy reduces cheating, but at the cost of increasing shirking.

The discussion so far is based on Ouchi (1979) and Eisenhardt (1985), but with the following differences. Ouchi (1979) argues that a minimum of internalized commitment is necessary in firms, for employees must accept the boss’s legitimate authority. I argue instead that bosses obtain employee acceptance of their authority by breaking the link between employee pay and performance. Hence acceptance of authority comes at the cost of shirking. Like Eisenhardt (1985), my approach puts more emphasis on rewards and on costs than that of Ouchi. But while she looks at alternative modes of control within the firm, I extend her analysis to a broader context which includes both markets and firms. For Eisenhardt, the choice between rewarding outcomes and rewarding behavior hinges on the differential cost of measuring each plus, in the case of outcome rewards, the cost of transferring risk to the agent. This requires that agents be risk averse. I make the more general assumption that both principal and agents are risk-neutral. Hence outcome uncertainty plays no role in my model. Lastly Eisenhardt (1985, p. 137) argues that:

If there is no outcome uncertainty or agents are not risk averse, then the choice between behavior and outcome control reduces to simply their comparative measurement costs.

By contrast, I posit that the cost of each organizing method is the sum of measurement costs and of the costs of the unwanted consequence of imperfect measurement. Hence the cost of replacing behavior by price (outcome) control is the cost of measuring outputs plus the negative side effects that arise from imperfectly measuring outputs, i.e., the reintroduction of cheating costs. This point is further developed in the following section.

2.4. The Choice between Firms and Markets

The argument made so far is summarized in Figure 1. Total organizing costs (CS') are the sum of cheating costs (line CC') and shirking costs (line SS'). (Recall that shirking and cheating costs are the sum of the costs of enforcement and the residual amount of loss; alternatively, they can be thought of as the total amount of loss sustained in the absence of enforcement.) Moving from the left to the right of the figure means replacing price constraints by behavior constraints. As behavior constraints replace price constraints, the level of cheating costs (CC') falls and the level of shirking costs (SS') rises, and vice versa as price constraints replace behavior constraints. This reflects the hypothesis that increased shirking is a necessary consequence of reducing cheating, while reducing shirking automatically increases cheating. The institution chosen will be that for which total organizing costs, CS', the sum of cheating and shirking costs, is lowest. In the figure, H represents a pure hierarchical solution (zero price constraints, 100 percent behavior constraints). This institution

---

11Whether organization costs will be passed on to third parties (e.g., consumers) does not depend on the system of control used, but instead on the degree of competition in the final product market. Hence a firm which fails to reject poor quality inputs and incorporates them into outputs will absorb the costs in the form of reduced future sales if there are substitute products on the market (cf. GM’s eleven point decline in its share of the North American market between 1980 and 1989). Whether poor quality is due to poor “buy” practices or failure to control shirking by employees (inefficient “make”) should have no impact on the extent to which these high organization costs are visited on consumers or absorbed by the firm.
will be chosen if the reduction in cheating costs achieved by replacing all price constraints by behavior constraints exceeds the resulting increase in shirking cost (Figure 1(a)). Similarly, $M$ represents a pure price solution (100 percent price constraints, zero behavior constraints), and will be chosen in the opposite case (Figure 1(b)). The type of transactions found in markets is clustered at $M$, while those found in firms are at $H$. The level of shirking relative to cheating costs, and hence the institution chosen to govern the transaction, will vary from transaction to transaction. Some transactions, for which output is relatively easy to measure and the consequences of not measuring it are low, but for which behavior is difficult to direct and monitor, will be organized most efficiently through the market. For other transactions, it may be difficult to measure all the relevant dimensions of output, and failure to do so may have severe consequences, but behavior may be relatively easy to constrain. These activities will be organized in firms.
An example may clarify the argument. Marketing services can be purchased in the market from independent agents—called manufacturer representatives, or “reps”—or they can be obtained from salaried employees (the hierarchical solution). Employing salaried salespeople can involve high shirking costs because there is little correlation in selling between behavior and outcome, making it extremely difficult to determine what a salesperson must do to close a sale successfully. It is also costly to supervise outside salespeople, who must visit customers in a wide variety of locations.

One way to reduce shirking costs is to use reps. Reps are independent contractors from having to direct and monitor the behavior of salespersons, as reps are incited to use their own methods since they are rewarded in function of their performance (Anderson 1985). The drawbacks of this solution are as expected: not all aspects of a salesperson’s performance are equally easy to measure. Reps rewarded by commissions can be expected to maximize their income by minimizing those aspects of performance which demand their time and effort but are difficult to observe. Reps will minimize product demonstration, instruction, and after-sales service (Wilkins 1970) and will maximize their income by selling established products (as opposed to new ones) to smaller, regular accounts (as opposed to new and large accounts) (Anderson and Oliver 1987). They also may indulge in unethical practices, such as misrepresenting product characteristics or customer needs (Anderson 1988, Robertson and Anderson 1990). Using reps will therefore be efficient when simple measures of output (such as sales volume) are adequate indices of selling performance and while selling behavior is a poor guide to performance.

When the imposition of price constraints on reps results in high cheating costs, it may be necessary to replace most price constraints by behavior constraints. For example, contracts which require reps to promote new products or to adopt low-pressure, expertise-based sales tactics are difficult to draft and to enforce, as the output of reps is then difficult to measure in the short run. Switching to full behavior constraints (hiring employees to do the selling) will reduce organizing costs if their behavior is relatively easy to monitor, while outcomes are difficult to measure and the consequences of imperfect measurement are high (Anderson and Oliver 1987). Empirical studies by Anderson (1985) and John and Weitz (1989) support this view.

3. Hybrid Arrangements

Figure 1 assumes that there is a linear relationship between the level of constraints and the amount of shirking/cheating. If this is the case, imposing a mix of behavior and price constraints will never minimize total organizing costs. However, most actual institutional arrangements do combine both price and behavior constraints. Transactions commonly referred to as market arrangements are governed mainly by prices but often include behavior constraints. Borrowing money, for example, is a market transaction. Yet it often involves the imposition of a significant number of behavior constraints: bond covenants usually limit dividend payments, future debt issues, and specify minimum levels of working capital (Jensen and Meckling 1976). Market transfer of knowledge through licensing is usually accompanied by various restrictions on the behavior of the technology buyer, curtailing his rights to export the product manufactured under license (Caves, Crookel and Killing 1983).

12 Salaried salespersons are rewarded on the basis of their behavior. The five most common criteria by which sales managers evaluated their employees were (in descending order): attitude, product knowledge, selling skills, appearance and manners, communication skills and sales volume (tied) (Jackson, Keith and Schlacter 1983).
Similarly, employment relationships which consist essentially of directives specifying behavior sometimes use prices to indirectly control some aspect of that behavior. About one-quarter of all workers in U.S. manufacturing industries in the mid-1970s (and 23 percent of all farm labor in 1959) were not remunerated solely in function of their obedience to managerial directives, but were paid, at least in part, in relation to their marginal product (Seiler 1984). Piece work schemes, bonuses paid to employees, and stock option plans, are examples of this type of arrangement. Such pay schemes are especially widespread at upper levels: bonuses made up 31 percent of the total compensation received by executive vice-presidents in 1986 (Reibstein 1987).

Figure 2 shows what happens when the level of cheating and shirking costs increases nonlinearly as one specializes into one method of organizing. The minimum of total organizing costs (CS') is then attained in institutions which combine price and
behavior constraints. This is because it becomes prohibitively costly to use either pure hierarchical or price methods. In that case, it will pay to combine the two organizing methods. Hence the use of behavior constraints in what are still thought as market transactions, and the use of price constraints within firms.

To return to our selling example, consider the case in which the costs of monitoring behavior are very high, while the deleterious side effects of using reps (such as the reps' reluctance to learn to demonstrate the product and to sell to new accounts) are few in number and relatively easy to control. It may be advisable, then, to use reps, but to impose on them a limited number of behavior constraints (point \( T \) in Figure 2b). For example, the contracts signed by British manufacturers with their overseas reps around the turn of the century—at a time where high communication costs made supervision at a distance very costly—stipulated minimum input requirements covering the amount of traveling, advertising, and showing the rep was required to do. Contracts also fixed the level of stocks to be held by the rep, required reps to employ engineers or salesmen with technical knowledge of the manufacturer's products, and gave the manufacturer the right to refuse orders, to monitor stocks, and to inspect the sales records (Nicholas 1983). There are obviously limits to this strategy, as reps will balk at contracts specifying in detail their behavior while they alone bear the pecuniary consequences. Therefore, they will accept these rules only if employers share in the consequences. They will demand that their compensation include a base salary (point \( T \) in Figure 2).

When the unwelcome side effects of using reps are significant, a switch to behavior control is indicated, but this unfortunately results in shirking. One way to keep shirking to acceptable levels is to reintroduce price controls and to pay salespeople a salary plus a commission related to the volume of their sales. Point \( B \) in Figure 2(a) represents such a compensation package in which commissions account for a small part of total pay. Hence if constraining behavior and measuring output both exhibit diminishing returns, most compensation plans will incorporate price and behavior constraints.

Why does using both sets of constraints simultaneously minimize total organizing costs? A salesperson's incentive to shirk under full behavior constraints is very strong, since he or she gets 100 percent of the time or effort thus saved (minus penalties if discovered). The same goes for a rep's incentives to cheat under a regime of pure price constraints. When the income of salespeople is partly taken in commissions, and partly in the form of salary, they are made to share in the cost of shirking and cheating. If they shirk, they make less commission; hence incentives to shirk are lower, although not as low as if they were under straight commission. Their incentive to cheat is lower than if they were under straight commission, since commissions now only make up part of their income. Hence the combined incentives to be opportunistic are lower if salespeople are simultaneously under both constraints. This benefit must be balanced against the offsetting costs of having to establish two sets of constraints (Barzel 1989). In some cases, but not in all, the savings in organizing costs more than make up for this additional cost, and the most efficient institutional arrangement is in the swollen middle.

This is the case in selling. Most salespeople's compensation plans do combine salary and commission. In 1989, about 70 percent of the manufacturing firms surveyed

---

13There is a parallel with taxes. The probability that individuals will cheat on their taxes rises more than proportionately with an increase in the tax rate. Hence one can reduce cheating (at equivalent tax revenues) by replacing a high rate tax with two taxes at low rates.

14Eisenhardt (1985) found that small stores could not afford to set up commission schemes for retail salespeople.
by the National Institute of Business Management reported paying some form of salary plus commission, with 14 percent playing straight salary and 19 percent straight commission (Sales and Marketing Management 1989).\textsuperscript{15} Furthermore, empirical studies show that the share of commission in an employee total compensation package increases the easier it is to measure output, the costlier it is to monitor behavior, and the less important are demonstration and service, two dimensions of performance which are difficult to measure (John and Weitz 1989).

4. Limits to the Use of Price Incentives in Firms

One implication of the model is that firms introduce price incentives to reduce the high cost of constraining employee behavior. The higher that cost, the more likely that firms will rely on these incentives. However, using price incentives is not a panacea, for it reintroduces cheating costs.

Under the hierarchical method of organizing, rewarding workers for following managerial directives shifts to the employer the task of knowing how workers are to perform their tasks. Employees are less likely to collect and volunteer information relevant to that process because it does not benefit them directly. Employers must also deduce performance from behavior: opportunities for shirking, and hence the use of price incentives in firms, increase if observed behavior does not provide good clues about the quality and/or the quantity of work performed (Eisenhardt 1985). The level of managerial capabilities should also influence the use of price incentives in firms. Less competent employers will have to devote more resources to learning what employees must do and to monitoring their behavior, ceteris paribus, than more competent employers. One would expect greater use of price incentives when management incurs high costs in gathering information on workers' production function and in monitoring their behavior than when management can perform these tasks relatively cheaply.

This brief argument indicates the following: (1) The use of price incentives within firms should be relatively greater the lower the level of managerial expertise. Improvements in managerial techniques should reduce their use, ceteris paribus. (2) At any point in time, large and diversified firms should make greater use of price incentives than small and nondiversified ones, (3) Employers will use price incentives for those activities about which they have limited knowledge and/or those which are costly to supervise.

Price incentives take two main forms in firms. Under piecework and commission schemes, part or all of the pay of a worker is linked to his output. The compensation of managers may also be tied to the performance of their subordinates. This requires that their units be organized as profit centers. The sections that follow discuss these options in detail.

4.1. Piecwork

Piecwork saves on monitoring costs; it frees management from having to know how workers can best perform their tasks and from monitoring their behavior. Piecwork harnesses the idiosyncratic knowledge that workers may have on how to perform their job most efficiently (Burawoy 1979). Whenever it is more costly for employers than for employees to obtain that knowledge, workers paid by the piece should achieve higher output than those paid on salary, and piecwork schemes will be implemented. Clark (1984) cites the results of a number of studies which compare

\textsuperscript{15}As expected, these percentages varied across sectors. For example, the proportion of straight commissions was higher in financial services.
the hourly rates of pieceworkers vs. hourly workers in a number of different occupations. Overall, pieceworkers earned between 13 and 25 percent more than hourly workers. Petersen's (1991) careful study of pay in the nonferrous foundry industry finds that, after controlling for establishment size, production technology, region, union coverage, occupation, and sex, piece rate workers earn 5 to 12 percent more per hour than time-rated workers.16

Piecework and its variants are found when entrepreneurs experience high costs constraining the behavior of their employees. This is the case when activities are geographically dispersed, such as in sales (as seen above), in agriculture, and in construction (Eccles 1981). Piecework and its associated price incentives were commonplace in the 19th century, before the development of sophisticated management techniques. Under the system of “inside contracting” a capitalist provided floor space, raw materials and machinery to a subcontractor and paid him a piece rate for his output. These subcontractors in turn hired their own employees, and trained, monitored, and paid them (Buttrick 1952). In mining, inside contracting was known as the “butty” or the “tribute” system (Jenkin 1948).

The reasons for the existence and the demise of inside contracting are consistent with our model. Inside contracting was chosen because “it supplied a ‘self-acting stimulus’ which dispensed with the necessity of incessant supervision of the managing foreman by the employer” (Taylor 1960, p. 216). In short, it was “a method of evading management” (Pollard 1965, p. 38). As we would expect, the maximization of effort by inside contractors led them to a free ride on the dimensions of performance which were not priced. Workers abused machinery and wasted materials and work-in-progress inventories (Buttrick 1952, Williamson 1975). In mining, there was a tendency to work only the best quality ore (Hillman 1984).

Under the influence of “scientific management,” inside contracting was displaced at the turn of the century by a system of “straight” piecework where wage-earning specialist supervisors set output norms and piece prices. This system was in turn replaced by hourly pay in the 1920s (Gospel 1983). In tin mining, the development of scientific prospecting methods made it unnecessary to rely on the miner's tacit knowledge of where to find veins, while the replacement of the pick by the electrical rock drill—a machine too costly to be abused—made close supervision cost effective. Both factors sealed the fate of inside contracting in tin mining (Burke 1982).17

4.2. Profit Centers

An examination of benefits and costs of hierarchical and price-based methods of organizing can also explain why firms choose to set up profit centers. As firms grow in scope and geographical reach, the cost of constraining employee behavior increases. Product diversification reduces the technical knowledge that headquarters (HQ) has of the production function used by subunits. Geographical dispersion, especially if it involves foreign countries, increases the difficulty of evaluating the performance of subunit managers, because both distance and variations in local conditions make interpretation of behavior more difficult.

Using price incentives can help alleviate this problem. Price incentives can be used at the level of the subunit by (1) separating the firm into “quasi firms” (profit centers)

---

16This cannot be due to self-selection, since firms using time rates have the possibility to fire least efficient workers and to keep the most efficient ones. Petersen (1991) also shows that risk-sharing accounts for about 20 percent of the average wage difference between piece-rated and time-rated workers. Thus the main task of piece rate schemes is not to share risk, but, as argued here, to induce effort.

17Because inside contracting and piecework was used to tap the worker's idiosyncratic knowledge, their replacement by hourly pay often involved a “deskilling” of the relevant jobs. I am indebted to an anonymous referee for bringing this point to my attention.
which buy their inputs and sell their outputs at arm’s length to other subunits; (2) giving full freedom to the subunit manager to maximize the profit of the subunit, and (3) rewarding subunit managers as a function of these results. This solution has been called the multidivisional (or M-form) organizational structure (Chandler 1962, Williamson 1975).

What benefits can the M-form structure offer? In situations where subunit managers know better than HQ how to maximize profits, rewarding them in proportion to these profits and giving them free rein to achieve these profits will yield higher profits than directing their actions from HQ. Subunit managers will be incited to make use of their specialized knowledge for their own benefit and therefore for the firm’s benefit. If all internal interactions are correctly priced, the firm will elicit from its subunit managers exactly the same behavior that it would obtain through perfect behavior control, but with zero monitoring costs. Hence, the higher the cost of monitoring subunit managers, the greater the benefits of M-form structures. Since both product and geographical diversity increase monitoring costs, diversified firms should benefit more from M-form structures than nondiversified ones. Chandler’s study (1962) of the adoption of the M-form supports this point. Fligstein (1985) found that product diversity was a significant determinant of the early adoption of the M-form by US corporations (although later adoption was also influenced by mimetism).18

The practical problems of setting up profit centers provide a good illustration of the unwelcome side effects of using price incentives in firms. To maximize their own income, managers of subunits will maximize the yearly profits made by their units. In the process, they will maximize the use of unpriced (or undervalued) inputs and minimize that of unpriced (or undervalued) outputs. For example, if HQ does not assign a price to a firm’s loss of reputation which results from a subunit’s unethical behavior, then subunit managers can be expected to engage in profitable but shady business practices. To avoid such unanticipated and undesirable side effects, all inputs and outputs used and produced by the profit center—including intangibles such as reputation and experience—need to be priced to reflect their cost and benefit to the firm as a whole. This is an impossible task (Ouchi 1979). The firm owes its very existence to the fact that some interactions between its parts cannot be organized through a price system. If all interactions could be priced, there would be no benefits to organizing within the firm. The number and relative importance of nonpriceable interactions will vary across firms and even within firms. The proportion of nonpriceable interactions is expected to be high in vertically integrated firms and in those whose subunits receive continuous innovations from a centralized R&D or market research facility. In each of these cases the market for the relevant inputs (tacit knowledge, reputation, or raw materials) is characterized by high transaction costs, making some form of hierarchy the preferred organizing method (Hennart 1982, 1991). Supplementing hierarchy with intrafirm markets to transfer such inputs and outputs between subunits generates pricing problems and results in suboptimization.19

---

18 Mahoney’s (1990) findings that the adoption of the M-form by US firms was influenced by their geographic dispersion and product diversification is also supportive of the argument, although his research design does not make it possible to separate the influence of this variable from that of mimetism.

19 Intertemporal suboptimization is also a problem. Subunits are not freestanding entities, quoted on local stock markets. There is therefore no easy way to evaluate the impact of the subsidiary manager’s present decisions on the subsidiary’s future profit stream. If the manager is rewarded on the basis of annual profits, he can be expected to maximize present income at the expense of future profits by a variety of stratagems, such as cutting R&D budgets, or cutting employment and jeopardizing long-term government relations.
To check the generation of unwanted side effects and to encourage desirable ones, HQ may supplement transfer prices with behavior constraints. The subunit manager will be told to maximize profits; in addition, he will be asked to transfer fixed quantities of inputs or outputs at fixed prices to other subunits, obey HQ directives on ethical behavior, worker safety, pollution control, and the like. This represents a shift toward hierarchy, and it reduces the advantages of profit centers. As the autonomy of subunit managers declines, so will their incentive to work hard and show initiative. HQ will now have to spend more resources directing and monitoring the subunit manager’s behavior. Both the informational simplicity and the motivational virtues of profit centers will be reduced. Hence the mix of price and behavior constraints faced by profit center managers should vary in direct proportion to the difficulty of pricing internal transactions. When internal transfers of inputs and outputs are few and can be priced easily, profit center managers should have considerable freedom to buy or sell from whomever they like at whatever price they may negotiate. In the words of Eccles (1985), they should have “full exchange autonomy.” Conversely, whenever internal transactions are numerous and difficult to price, HQ will dictate from whom managers may buy, to whom they may sell, and the prices for these transactions. There are fewer exchanges of raw materials, reputation, and know-how in diversified firms than in single-business, vertically integrated concerns, and hence profit center managers should have greater autonomy in the former than in the latter.

Studies of profit centers in firms support this view. Eccles (1985, p. 114) found that 39 percent of the heads of profit centers of the US firms he studied had full exchange autonomy. As expected, that percentage was higher (50.5 percent) for highly diversified firms (unrelated businesses) than for single business, vertically integrated firms (29 percent). Another way of evaluating how much exchange autonomy profit centers have (and hence of ascertaining the degree to which price incentives are used) is to look at the type of transfer pricing used in transactions between profit centers. Transfer at market prices is dysfunctional if there are significant nonpriceable exchanges between subunits. Hill, Hitt, and Hoskisson (1990) found that highly diversified firms were more likely to use market-based transfer prices for internal transactions than undiversified ones. This supports the view that the use of profit center in firms is limited by cheating costs.20

Earlier we have argued that using price incentives reduces shirking but necessarily increases cheating. Reintroducing price incentives in firms can take the form of piecework schemes and profit centers. Both of these techniques suffer from an unwelcome side effect: the reappearance of cheating costs. The evidence shows that piecework and profit centers tend to be used whenever they significantly reduce shirking costs with only a minimum increase in cheating costs.

5. Conclusion

The preceding pages have developed a “comparative institutional” model in which economic institutions are analyzed as a mix of two pure organizing methods, hierarchy and the price system. The model shows that prices and hierarchy differ in efficiency because they use different methods to organize activities, and that the comparative advantage that firms have over markets in organizing certain transactions derives from their greater reliance on hierarchy. Hence the comparative institutional model answers Demsetz’s (1988, p. 149) call for a theory which clearly differentiates between the cost of transacting and that of managing. In the comparative institutional

---

20 Only half of all transfers were at market or negotiated prices.
model, the cost of transacting is the cost of measuring output in all of its dimensions and the consequence of not measuring it perfectly; the cost of managing is the cost of directing and observing behavior and of failing to do it perfectly.

The comparative institutional model defines firms as institutions which rely primarily on behavior constraints, and markets as institutions which use mostly price incentives. No special theory is needed to account for hybrid institutions; they can be analyzed as combinations of the two organizing methods of price and hierarchy. For example, the model locates contracts in the continuum between market and firms, and defines them as consisting of limited behavior constraints within dominant price constraints. The concentration of institutional forms in the swollen middle between markets and firms can be explained by increasing organizing costs as one specializes into using either pure price or pure hierarchical methods. In other words, organizing costs are often lower when individuals are partially constrained by both control systems simultaneously than when they are subject to full behavior or price constraints. Hybrids will be used when these savings from lower organization costs more than make up for the additional cost of imposing two sets of constraints.

The model also clarifies the relationship between hierarchy and firms. By reviewing the problems incurred in using price incentives within firms, it shows that the use of hierarchy is a sine qua non condition for the existence of firms. It is not because they mimic the price system that firms can sometimes incur lower organizing costs than markets, but because they replace price constraints with behavior constraints. In some cases the high cost of establishing a pure system of behavior constraints may lead firms to selectively reintroduce some degree of price constraints, but behavior constraints will still be the firm’s primary organizing method.

Lastly, while there are clear conceptual differences between prices and hierarchy as methods of organizing, the distinction between “firm” and “market” is sometimes ambiguous; so-called “market” transactions may take on many “firm” characteristics while “firm” arrangements may have “market” overtones. But this ambiguity does not invalidate the usefulness of distinguishing between markets and firms if a clear distinction is made between organizing methods (prices and hierarchy) and institutions (markets and firms). The costs and benefits of any institution can then be analyzed in terms of its relative mix of the clearly defined and conceptually distinct pure categories of “price” and “hierarchy.”

Acknowledgements

Earlier versions of this paper were presented at the conference on Organizational Theory and the Multinational Corporation at INSEAD in September 1989 and at the Academy of International Business meetings in Toronto in October 1990. I thank Arie Lewin, Arthur Stinchcombe, and the anonymous referees, as well as Erin Anderson, Yves Doz, Moshe Farjoun, Gunnar Hedlund, John Kimberly, Steve Kobrin, Guido Krickx, Huseyin Leblebici, Joseph Mahoney, Young-Ryeol Park, Thomas Roehl, Ming-Je Tang, Jose de la Torre, Andy Van de Ven, Keith Weigelt, Sidney Winter, Chow-Ming Joseph Yu, and especially Kathleen Saul, for their helpful comments.

References


