The Firm as a Subeconomy

Bengt Holmstrom
Massachusetts Institute of Technology

This article explores the economic role of the firm in a market economy. The analysis begins with a discussion and critique of the property rights approach to the theory of the firm as exposited in the recent work by Hart and Moore ("Property Rights and the Nature of the Firm"). It is argued that the Hart–Moore model, taken literally, can only explain why individuals own assets, but not why firms own assets. In particular, the logic of the model suggests that each asset should be free standing in order to provide maximal flexibility for the design of individual incentives. These implications run counter to fact. One of the key features of the modern firm is that it owns essentially all the productive assets that it employs. Employees rarely own any assets; they only contribute human capital. Why is the ownership of assets clustered in firms? This article outlines an answer based on the notion that control over physical assets gives control over contracting rights to those assets. Metaphorically, the firm is viewed as a miniature economy, an "island" economy, in which asset ownership conveys the CEO the power to define the "rules of the game," that is, the ability to restructure the incentives of those that accept to do business on (or with) the island. The desire to regulate trade in this fashion stems from contractual externalities characteristic of imperfect information environments. The inability to regulate all trade through a single firm stems from the value of exit rights as an incentive instrument and a tool to discipline the abuse of power.

First version presented at the conference, "The Theory of the Firm and Economic Transition," Beijing, China, September 1–2, 1996. This version presented at the conference, "Bureaucracy: Issues and Apparatus," Marconi Center, Marshal, California, April 4–5, 1998. Portions of this article stem from my third Yrjö Jahnsson Lecture, delivered in November 1992. The research was partially funded by the National Science Foundation. I want to thank Oliver Hart, David Kreps, Paul Milgrom, John Roberts, Jean Tirole, and Oliver Williamson for many fruitful conversations on this topic and Abhijit Banerjee, Tore Ellington, Guido Friebel, Robert Gibbons, Benjamin Hermalin, Peter Högfeldt, Jon Levin, Lars Stole, Michael Waldman, and Robin Wells for comments on earlier drafts.

1. Introduction

Coase's (1937) key insight was to observe that in organizing transactions, there is a choice between placing a transaction in a market and locating it inside a firm, and that therefore this choice must have a bearing on understanding both forms of trade. It took a long time for Coase's insight to sink in, but today, thanks to the efforts of Arrow, Alchian, Demsetz, Williamson, and many others, institutional theory has again found a major place in economic theorizing. Initially theorists focused on problems of contracting under asymmetric information, developing new methods to deal with the strategic use of information and the attendant incentive problems. Equipped with insights from basic contract theory, theorists have again returned to the classic question raised by Coase: What determines the boundaries of the firm? The economic significance of this question is easy to appreciate today as mergers and acquisition activity has reached unprecedented levels in this past decade.[1]

The new theories remain scattered and it would take a long article to survey all approaches and strands. That is not my intent. Rather my objective is to suggest a partial synthesis of the theory of the firm, integrating aspects of internal organization with the leading modeling approach to boundaries, the modern property rights theory. In the envisioned synthesis, measurement costs of the kind originally stressed by Alchian and Demsetz will return to center stage. For this reason, I will start out reviewing Alchian and Demsetz's (1972) vision of the firm as a monitor of inputs. Their model will also serve as an important backdrop for interpreting the influential work of Grossman, Hart, and Moore on incomplete contracting and asset ownership. Incomplete contract models stress the importance of asset ownership as an instrument for influencing individual incentives. Ownership both creates and resolves holdup problems, altering the returns to investment in human capital. The most advanced exposition of this line of analysis is found in Hart and Moore (1990). I provide a sketch of the Hart–Moore model in Section 3, followed by a critique. I argue that this model, despite its express objective to explain the boundaries of firms, fails to do so, at least if the model is interpreted literally. The model offers a theory of individual ownership of assets, that is, how control over assets should be distributed among individuals, but it does not explain why firms own assets, let alone why firms tend to own or have control over all the alienable, nonhuman assets that they use to produce goods or in other ways serve customers. I take the ownership of assets by firms, and the attendant feature that economic contracts are made with firms, not with their employees or owners, as one of the most significant and robust empirical regularities to be explained by any theory of the firm.[2]

Having said this let me confess at the start that I am unable myself to offer a well-developed explanation of asset ownership by firms. But I have a conjecture that is inspired by the Hart–Moore article. The conjecture is that the firm gains

1. Transactions worth over $900 billion occurred in 1997 in the United States.

2. The "nexus of contracts" view, discussed in Alchian and Demsetz (1972), suggests as an explanation the costs of writing bilateral contracts between parties in a network of relationships, but I have not found a coherent explanation of why these costs are any less within firms.

power over human capital through ownership and control of assets. Ownership confers contracting rights that allow the firm to decide who should be offered the opportunity to work with particular assets and on what terms.[3] The firm uses its rights to set "internal rules of the game" and to design incentives in a manner that internalizes some of the contractual externalities that are present in markets due to asymmetric information. The hold-up problem is just one example of a contractual externality. There are several others, as I will illustrate.

In effect, I see the firm as a subeconomy in which the executive office has the power to regulate trade by assigning tasks, delegating authority, and delineating principles for how explicit and implicit incentives are to be structured. There is an obvious analogy between my vision of the firm as a regulator of trade and the role that a government plays in determining tax and other rules for an economy. It is interesting to ask which tasks should be left for the government given that firms can internalize a lot of externalities within their local trading environments. I will not pursue that question here, except to note that the principles of second-best, which commonly have been used in the context of public finance, seem to have much greater applicability and relevance in the context of the theory of the firm. Of importance, the second-best logic can explain why firms make extensive use of low-powered incentives and narrow job designs: to encourage employee cooperation in situations where incentives for cooperation would be compromised by excessively strong market incentives. In this view the much-lamented lack of individual initiative in bureaucratic firms should be seen as an essential element in providing incentives for cooperative activities. Section 4 illustrates this idea with a few simple examples, which are variations on my earlier work with Milgrom on multitask agency (Holmstrom and Milgrom, 1991, 1994; see also Laffont and Tirole, 1991).

Section 5 turns to the question of boundaries, applying the examples from Section 4 to the multifirm case. Section 6 compares the subeconomy view with the property rights view. The predictions are empirically distinguishable, because in the subeconomy view measurement costs play a central role in explaining contractual externalities and the need to internalize them. More importantly, though, the subeconomy view suggests that the boundary question is of interest, not so much in isolation as in the context of how firm boundaries condition what kinds of incentives can be structured on markets as well as within firms. The strength of the property rights view is that it articulates so clearly the role of market incentives and how they can be altered by shifts in asset ownership. But it says nothing about the incentives that can be created within firms. The real challenge is to understand how the two forms of organization complement each other as well as compete with each other as mechanisms for

3. Models of incomplete contracting have almost exclusively focused on the ex post bargainning rights and implied residual returns that come with ownership. My emphasis is on ex ante contracting rights. Tirole (1994) also stresses that ownership entails broader contracting rights than typically considered in the property rights theory. Rajan and Zingales (1998) emphasize the fact that the owner controls employee access to assets, human as well as physical. This is very similar to my approach, which is based on Holmstrom and Milgrom (1991, 1994).

influencing individual incentives. The main objective of this article is to point research in that direction.

2. Alchian and Demsetz Revisited

2.1 The Free-Rider Problem

The earliest theories by Coase (1937) and Simon (1951) suggested that the distinguishing characteristic of a firm is its use of authority. In firms, resources are allocated by fiat rather than by prices. Alchian and Demsetz (1972) objected to this interpretation by noting that market participants seem to have similar authority by virtue of their right to withdraw from trade. An individual who stops buying from her grocer is just as much "firing" her grocer as an employer who terminates an employee. The context and consequences are different, but the economic mechanism is the same. Instead of authority, they suggested that the key characteristic of a firm is the monitoring of inputs rather than outputs. The need for monitoring arises, because much output is jointly produced so that individual contributions get disguised by the contributions of other parties. This results in free riding and a suboptimal level of performance. It is useful to sketch a formal model of their argument, since it will facilitate the interpretation of Hart and Moore's model.[4]

Suppose there are n workers supplying inputs into a production process that yields a joint output $y = y(e_1, \ldots, e_n)$, where e_i is the input of worker i. Assume each e_i, referred to generically as effort, is unobservable and represents the private cost to worker i of choosing effort level e_i. Only output y is observable. As such, y can be used as a basis for a contract between the workers. Let $s_i(y)$ represent worker i's share of output in the contract and assume for the moment that shares are chosen so that $\Sigma_i s_i(y) = y$, that is, total output is exactly distributed among the workers, because workers form a partnership. The question then arises: Can the sharing rules s_i be chosen so that they induce workers to choose inputs efficiently? If we take workers to be risk neutral and assume that they have sufficient wealth, an efficient vector of inputs e^* is found by maximizing total surplus. Assuming a differentiable, strictly concave production function, the following first-order conditions fully characterize e^*:

$$\frac{\partial y(e^*)}{\partial e_i} = 1, \text{ for all } i = 1, \ldots, n. \tag{1}$$

Since inputs are privately chosen, it is natural to assume that they are chosen noncooperatively with a Nash equilibrium characterizing the equilibrium outcome. Assuming that sharing rules are differentiable, an assumption that simplifies the exposition, but is inconsequential for the conclusion, we have the following first-order conditions for a Nash equilibrium e^{NE}:

$$\frac{ds_i(y(e^{NE}))}{dy}\frac{\partial y(e^{NE})}{\partial e_i} = 1, \text{ for all } i = 1, \ldots, n. \tag{2}$$

4. The model is from Holmstrom (1982).

If we are to have $e^{NE} = e^*$, we must have

$$\frac{ds_i(y(e^*))}{dy} = 1, \text{ for all } i = 1, \ldots, n. \tag{3}$$

This is just the standard result that efficiency requires each worker to obtain the marginal social surplus from his actions. On the other hand, since sharing rules have to add up to the output level y no matter what y is, we see, by differentiating the budget constraint, that

$$\sum_i \frac{ds_i(y(e^*))}{dy} = 1. \tag{4}$$

Evidently, Equations (3) and (4) are inconsistent, implying that a Nash equilibrium e^{NE} can never equal the efficient choice e^*. The problem is that there are not enough margins to go around. By Equation (3), each worker should receive 100% of his marginal product, but the budget offers only a total margin of 100% to be divided among the workers. The split of the total margin leads to a suboptimal choice of inputs.

It is worth emphasizing that the free-rider problem has nothing to do with externalities in the production function (as argued by Alchian and Demsetz). As shown above, the problem arises just as easily with an additive production function in which there are no interactions between the workers' inputs. The free-rider problem is exclusively driven by informational externalities caused by the inability to identify individual contributions.[5]

2.2 Two Solutions to the Free-Rider Problem

There are two solutions to the free-rider problem. The first one is to allow for budget-breaking sharing rules that do not add up to the total output at all levels of y (Holmstrom, 1982). The second is to expand on measures through additional monitoring. Let me start with the budget-breaking solution. If we allow sharing rules satisfying the weaker constraint $\Sigma_i s_i(y) \leq y$ for all y, then efficiency can be restored. All one has to do is to let every worker suffer sufficiently if output falls below the efficient level $y(e^*)$. For instance, let $s_i(y) = \frac{y}{n}$, if $y \geq y(e^*)$ and $s_i(y) = \frac{y(e^*)}{n} + y - y(e^*)$, if $y < y(e^*)$. This rule burdens each agent with the full social cost of shirking. It breaks the budget by requiring the partnership to throw away some of the output if the efficient output level is not reached. But since the rule will induce workers to choose the efficient level of effort, no output needs to be thrown away in equilibrium. There is a question, of course, whether such a self-imposed threat is credible. To deal with potential credibility problems, one can bring in a passive outsider that breaks the budget. This has its own problems with collusion, but I will not go further into that issue here.

Alchian and Demsetz (1972) suggested that the free-rider problem be solved by introducing a monitor to assess worker inputs. A monitor, who is able to

5. That technological externalities as such don't determine firm boundaries is evident. Just think of a car or any other complex product, in which almost all components are tightly engineered to fit, yet hundreds of firms are involved in the production process.

perfectly discern individual inputs, can use a simple forcing contract to induce the right choice of effort. In such a contract, the worker is paid only if he performs to the standard. This suggestion leaves two questions unanswered: What assures that the monitor doesn't cheat on payments when performance meets the standard? And what gives the monitor the incentive to monitor? The latter question, according to Alchian and Demsetz, is resolved by giving the monitor the claim on the residual, that is, the output net of payments to the workers. In their interpretation, this arrangement makes the monitor also the owner. The first question they never address, but one can imagine that in a repeated game, the owner's concern for her reputation will prevent cheating [see Bull (1987) and Baker, Gibbons, and Murphy (1997)]. I will not pursue this point either. Instead, let me look at the question of monitoring from a more abstract perspective, namely as a method to create a richer system of performance measures.

2.3 Additional Performance Measures

When will an additional performance measure, say $z = z(e_1, \ldots, e_n)$, add value by enabling a better Nash equilibrium? To make the analysis more transparent, assume that there are just two workers whose joint output y is additively separable. With an appropriate transformation of variables, we can write $y = e_1 + e_2$, leaving the workers with strictly convex cost functions $c_1(e_1)$ and $c_2(e_2)$. Let $z = e_1 + \gamma e_2$ be the additional performance measure with γ a constant. Note that we can restrict attention to linear sharing rules, $s_1(y, z) = \alpha y + \beta z + \delta$ and $s_2(y, z) = (1 - \alpha)y - \beta z - \delta$. Any set of sharing rules that implements a particular vector of efforts (e_1,e_2) can be replaced by a set of linear sharing rules with weights equal to the derivatives of the general rules evaluated at the implemented vector of efforts. Because of budget balancing, the weights of y have to add to one and the weights of z have to add to zero. Noting this, the workers' choice of efforts will be described by the first-order conditions

$$\begin{aligned} \alpha + \beta &= c_1'(e_1) \\ (1 - \alpha) - \beta\gamma &= c_2'(e_2). \end{aligned} \tag{5}$$

According to Equation (5), the workers will respond to changes in β as follows:

$$\begin{aligned} \frac{de_1}{d\beta} &= \frac{1}{c_1''} \\ \frac{de_2}{d\beta} &= \frac{-\gamma}{c_2''}. \end{aligned}$$

The effect of a change in β on net output, $y - c_1 - c_2$, can after a bit of algebra be expressed as

$$\left[(1 - \alpha)\frac{1}{c_1''} - \alpha\frac{\gamma}{c_2''}\right] d\beta. \tag{6}$$

Evidently the additional measure z will be valuable if and only if the bracketed

term is nonzero, since then we can choose a nonzero β that makes Equation (6) positive.

If $\gamma = 1$, then $z = y$. In this case, of course, the additional measure adds no value. Formally, this is true because changes in β are equivalent to changes in α. It follows that for the second-best value of α (the value that maximizes net output when y is the only available measure), the bracket in Equation (6) must be zero (I am assuming an interior solution, which is guaranteed if the cost functions have zero derivatives at zero). Consequently, as long as γ is different from 1, the expression in Equation (6) must be nonzero at the second-best value of α and hence net output can be raised either by increasing or decreasing β. Specifically, if $\gamma > 1$, then net profit is raised by some $\beta < 0$ and conversely if $\gamma < 1$. This is perfectly intuitive. When $\gamma > 1$, the z measure is more sensitive to changes in e_2 than to e_1 and hence worker 2 should be given a positive share of z, that is, β should be negative.

The economic point of this exercise is simple. As long as the additional measure z is not collinear with output y, it provides additional information about the inputs and such information can always be used to strengthen worker incentives somewhat. This can best be seen by adding up the incentive coefficients of the two workers, the terms on the left-hand side of the expressions in Equation (5). The sum is $1 + (1 - \gamma)\beta$, which is greater than 1, because we will set $\beta > 0$ (< 0) when $\gamma < 1$ (> 1). In contrast, when output y is the only measure, and budget balancing is imposed, the incentive coefficients necessarily sum up to one. The important effect of an additional measure is to "break the margin budget" so that more incentive margins can be distributed to the workers. This observation will have an interesting parallel in the Hart–Moore model.

The analysis above extends readily to n workers with separable output. In order for z to admit a marginal improvement in the equilibrium outcome, the vectors $[dy/de_1, \ldots, dy/de_n]$ and $[dz/de_1, \ldots, dz/de_n]$, evaluated at the second-best Nash equilibrium e^{EN}, must not be collinear.

3. Ownership and Incomplete Contracts

3.1 The Role of Incomplete Contracts

Alchian and Demsetz' idea of monitoring has been criticized because it fails to explain where the boundaries between firms are drawn. The argument is that the proposed solution is insensitive to the institutional context. Any contract between the owner-monitor and the workers could just as well be carried out in a market as within a firm. The organizational location of the parties is inconsequential, since the contract specifies all the relevant contingencies.[6]

When contracts are incomplete in the sense that they cannot incorporate all future contracting opportunities, governance becomes consequential. Contractual incompleteness is of course endemic, both because contingencies are too numerous to deal with and because many relevant contingencies cannot

6. This critique is most compelling if we think of the contract as explicitly written. Similar enforcement rules apply inside and outside the firm when it comes to human capital. The argument is weaker if we think of the contract as implicitly enforced, say through repeated interactions.

be foreseen. When unspecified contingencies occur, the continuation of the relationship must rest on negotiating how the old contract should be completed or negotiating an entirely new contract. The outcome of these negotiations will generally depend on the organizational affiliation of the involved parties. Two individuals will negotiate under very different incentives as employees of a single firm than as owners of separate firms, for instance.

The role of governance was first discussed in the context of vertical integration (Williamson, 1979; Klein, Crawford, and Alchian 1978). When there are two separate firms, each owning assets that are critical for production, either side can cause economic damage to the other by withholding use of its assets. The possibility of holdups can be troublesome if the relationship requires initial investments that cannot be fully recouped by the investing party when the relationship is severed. The mere threat of terminating the relationship may allow the noninvesting side to appropriate enough of the rents so that, even when the relationship is continued under a new agreement, the investing side will not be able to recover its initial investment outlays. Under such a threat, the investing party will rationally underinvest in the relationship.

Businessmen are acutely aware of the holdup problem. For instance, joint venture agreements pay careful attention to the potential for holdups (showing that holdup problems often are dealt with contractually rather than by integration; see Holmstrom and Roberts, 1998). Various contractual levers are given to each side to protect relationship-specific investments, such as exclusive selling rights, long-term delivery agreements, rights to veto expansions, and so on. Also, these agreements are quite explicit about exit rights: what to do in case of an impasse in negotiations over how to handle unstipulated events. Some variant of "Russian roulette" (also known as the "Texas Shoot Out") is commonly employed as a termination rule. The rule requires one side to propose a price, which the other side can accept either as a selling price or as a buying price of the joint venture. In both cases, the outcome is 100% ownership by one of the parties.

Applied to vertical integration, the logic of holdup suggests that when investments entail a high degree of asset specificity—that is, assets are much more valuable within a relationship than outside it—integration is preferred to having two separate firms. Williamson (1985) provides a detailed discussion of the role of asset specificity. He stresses that asset specificity, particularly in conjunction with environmental uncertainty and transaction frequency are conducive to integration. A key premise of his transaction cost economics is that one can predict integration based on the characteristics of individual transactions rather than bundles of transactions.

Hart (1995) has argued that Williamson's transaction cost approach does not explain exactly what changes when two firms integrate. Specifically, why is the holdup problem mitigated by integration and what costs are associated with the decision? Hart's own answer, first presented in his work with Grossman (Grossman and Hart, 1986), identifies changes in residual rights as the critical variable. Ownership, by legal definition, determines who has the residual right to decide how to use the asset. An owner can contract out some of the rights,

say by leasing the asset for a specified time, but apart from such contractually assigned specific decision rights, an owner is free to use the asset as he sees fit (subject to legal and social constraints, of course). Ownership gives formal control over the asset for uses that have not been preassigned—it gives "residual rights of control."

While ownership may save a firm from being held up by another firm, it also provides a lever to hold up other parties that make valuable investments. There is both a benefit and a cost to a switch in ownership. "Entrepreneurship" or "initiative" are terms we hear when the positive incentive effects of ownership dominate. "Holdup" is reserved for the negative effects. Which term is used varies with the context, but the basic mechanism is the same. The fact that the property rights theory applies one and the same logic to characterize both benefits and costs is a significant conceptual advantage. Another big advantage is that the idea can be readily translated into formal models, as evidenced by the many articles on incomplete contracting written in recent years. The most elegant exposition of the theory is offered by Hart and Moore (1990). Let me turn to a simplified version of their model.

3.2 The Hart–Moore Model

The model can handle many parties, but I will assume that there are only two, a buyer B and a seller S. There are two dates, 1 and 2. There is a set of alienable assets A, the ownership of which can be assigned in various configurations to B and S (or possibly some unrelated outsider O). At date 1 both the seller and the buyer make unobserved investments, denoted b and s, respectively. A critical assumption of the model is that these investments enhance the value of each party's human capital rather than the value of the assets directly. The value of the assets may go up indirectly, however, as there are interaction effects between assets and human capital.

It is assumed that no contract can be written at date 1 regarding the use of the assets at date 2, nor can a meaningful contract be written regarding trades at date 2. This, of course, is quite unrealistic, but it serves to highlight the role of asset ownership in the starkest way possible. Because there is no contract ex ante, the parties have to negotiate about what if any trade should take place at date 2. Rather than specify a particular production technology and the associated revenues and costs from trading, it is sometimes more illuminating to define the outcome of the process abstractly in terms of values that accrue to the parties at date 2 as a function of their investments and the assets that they control in the final agreement. There are two possible outcomes at date 2. Either the parties agree to trade, or they go their own ways. If they trade, they split the total surplus, denoted $v(b, s)$, in a fashion to be determined momentarily. Note that the notation $v(b, s)$ assumes that if an agreement is reached, all the assets will be under the control of B and S; it will become clear shortly that it is never desirable to introduce a third party, who might have ownership of some of the assets, so this notation is justified. If the parties fail to reach an agreement, the buyer receives his best alternative payoff $v_b(b|A_b)$ and the seller her outside option $v_s(s|A_s)$. Here $A_b(A_s)$ is the set of assets that the buyer (seller) has

control over at date 2. Note that v_b does not depend on the seller's investment s, because without the seller's participation her human capital cannot benefit B. Similarily for v_s. This is the import of assuming that investments are in human rather than nonhuman capital.

Assume that $v \geq v_b + v_s$. In that case it is always desirable to reach agreement and divide the value of v between B and S in a fashion that gives b at least v_b and the seller at least v_s. The predictions of the model are (unfortunately) sensitive to the particular bargaining assumptions used to determine how the surplus $v - v_b - v_s$ is split (see DeMeza and Lockwood, 1997). The standard bargaining assumption is a 50:50 split, in which case the payoffs to the parties at date 2 will be

$$\begin{aligned} s_b(b, s|A_b, A_s) &= v_b(b|A_b) + 1/2[v(b, s) - v_b(b|A_b) - v_s(s|A_s)] \\ &= 1/2v(b, s) + 1/2v_b(b|A_b) - 1/2v_s(s|A_s), \qquad (7) \\ s_s(b, s|A_b, A_s) &= 1/2v(b, s) - 1/2v_b(b|A_b) + 1/2v_s(s|A_s). \end{aligned}$$

The sum of the payoffs add up to v, of course. Foreseeing these date 2 payoffs, B and S make their investment decisions b and s at date 1. It is assumed that the choices will be made noncooperatively (since they cannot be observed) and that a Nash equilibrium will result.

The value functions v, v_b, and v_s are exogenous to the model. The only endogenous variable influencing the parties choice of investment is the allocation of assets A_b and A_s. Note from Equation (7) that it is only through a change in the outside options that the asset allocation matters.

In order to study the effects of ownership on investment decisions, one needs to make further assumptions about how the endogenous variables interact. Hart and Moore make the plausible assumption that with more assets under one's control, the marginal incentive to invest increases. That is, $\partial v_b/\partial b(\partial v_s/\partial s)$ increases as B (S) controls more assets. Also, they assume that b and s are complementary inputs in the joint production function $v(b, s)$; the cross-partial is positive. Finally, the marginal contributions of b and s are assumed to be highest when the parties work together: $\partial v/\partial b \geq \partial v_b/\partial b$ and correspondingly for s.[7]

The first implication is that the equilibrium level of investment is at or below the efficient level. Consider the decision of B, for instance. Starting from the efficient level, B's investment incentives will be driven by $1/2\partial v/\partial b + 1/2\partial v_b/\partial b \leq \partial v/\partial b$. If the inequality is strict, B will underinvest. But the more B underinvests, the lower are S's marginal returns from investing, because investments are complementary. Once S reduces her investment, B again lowers his and so it goes until a suboptimal equilibrium is reached. This is all a straightforward consequence of supermodularity, of course.

A second implication of supermodularity is that it never pays to have joint ownership. In joint ownership either side can veto the use of an asset.[8] That

7. In more technical jargon, these assumptions state that the value functions are supermodular, that is, the marginal return from increasing a variable increases with the other variables.

8. The assumption that both sides can veto the use of an asset may not represent reality. As I

means that in the event of disagreement, some assets will be made unavailable for use by either side. Looking at Equation (7), it is clear that it would always be better to reassign a jointly owned asset to sole ownership by one party, say B for concreteness, since that increases B's incentives to invest without directly affecting S's incentives. The fact that joint ownership is suboptimal also means that it is never desirable to give control over an asset to an outsider who makes no investments. Doing so would just reduce the incentives for B and S to invest for the same reasons as discussed above. In addition, the shares in the value functions allocated to B and S would go down (using Shapley value as the extended bargaining outcome), reducing investments still further.

Supermodularity further implies that assets that are complementary, in the strict sense that one asset is worthless without access to the other asset, should be owned by the same party. If the two assets were owned separately, this would be equivalent to joint ownership in which both sides can veto the use of the assets, an arrangement we have just seen to be suboptimal. Conversely, if the assets are unrelated (in the sense that no person's investment affects both assets), they should be owned separately.[9]

Finally, note that multidimensional investment decisions, for instance, B allocating his investment between enhancing the outside value v_b and enhancing the joint value v, will always bias investments excessively toward the outside option. This is a form of rent seeking, where total value is sacrificed in order to achieve a larger share of the common pie (Holmstrom and Tirole, 1991; Rajan and Zingales, 1998). I will return to the organizational implications of this case shortly.

3.3 A Comparison with Alchian and Demsetz

It is illuminating to compare the Hart–Moore model to the free-rider model discussed in the previous section. To this end, first observe that the two models deal with exactly the same type of problem. In the Hart–Moore model the joint output is $y = v(b, s)$ and the issue is how to provide the parties with incentives to supply unobserved inputs $e_1 = b$ and $e_2 = s$. The only difference between the models is in the instruments they use to motivate the agents. In the free-rider model it is assumed that the joint product can be split between the workers according to arbitrary sharing rules. In the Hart–Moore model, instead of general sharing rules, the parties are constrained to use sharing rules that arise out of varying asset allocations as described in Equation (7). At first, this may appear to offer much more limited choices. For instance, when there is a single asset that has to be fully allocated to one party or the other, only two sharing

mentioned earlier, irreconcilable differences will often lead to termination *and* a transfer of assets to one party.

9. One might think that supermodularity also implies that when the marginal value of B's investment, say, goes up in the total surplus function v, then B should be given no fewer assets than before. In other words, the more important B becomes as an investor, the more assets (weakly) B should control. This is not true in general and requires additional assumptions that I will not get into.

rules are available. However, if one allows ex post randomization so that asset ownership can be determined not just ex ante, but also ex post through an ex ante agreed upon lottery, the set of possible payoffs in the Hart–Moore model expands to include all linear combinations of Equation (7). In another respect, the incentives in Equation (7) may be richer than in the free-rider model. The payoffs in Equation (7) cannot in general be written as a function of the joint output v, because b and s enter not just through v but also through the outside options v_b and v_s.[10]

Let me elaborate on this important distinction. Alchian and Demsetz (1972) suggested that organizations can deal with the free-rider problem by *internal* monitoring, which as we saw, serves to break the "margin budget," that is, the constraint that margins cannot add up to more than 1. The Hart–Moore model highlights the fact that the market can also break the margin budget in a different way, through external monitoring. When the parties bargain over the split of the surplus v, it is implicitly assumed that they can observe each other's outside options v_b and v_s; else it would not be possible to write the payoffs in the form of Equation (7). Unless v_b and v_s are collinear with v, they then qualify as additional, valuable measures of performance as discussed earlier. This is well illustrated by considering the special case where, for some asset allocation (A_b, A_s) the outside options add up exactly to total surplus:

$$v(b, s) = v_b(b|A_b) + v_s(s|A_s), \text{ for all } b, s. \tag{8}$$

Since the outside options are as valuable as the inside options, it is natural to think of this case as representing a perfectly competitive market. Computing the bargaining payoffs according to Equation (7) one finds that $\partial s_b/\partial b = \partial v/\partial b$ and $\partial s_s/\partial s = \partial v/\partial s$, that is, both B's and S's marginal returns from investment coincide with the social return. This results in an efficient level of investment. While both sides get only 50% of the marginal social return through the first term in Equation (7), they pick up another 50% of the margin through the additional performance measures v_b and v_s. Market monitoring breaks the margin budget, allowing both sides to receive 100% of the social return.

I am trying to make two points here. The first is that bargaining generates information, and this feature is a central part of the Hart–Moore model. Indeed, the model captures very nicely the essence of market competition: the right to exit a relationship and the information and incentives that the ensuing bargaining process generates. The role of the market had always been somewhat of an enigma in organizational analyses prior to the property rights models. It was common to take the market as a first-best default with internal organization

10. One option that is unavailable in the Hart–Moore model is budget breaking of the kind that can prove effective in the free-rider model. While third-party ownership can reduce the stakes of the buyer and the seller, this will merely weaken incentives rather than strengthen them. The problem is that the third party cannot punish the workers by taking part ownership. Nor can the third party add more rewards, because ownership shares have to add up to 100%. However, with richer contract options, budget breaking may be achieved also in incomplete contract models; see Maskin and Tirole (1999).

emerging where the market failed (e.g., Williamson, 1975). That view did not explain exactly what makes markets so useful and how asset ownership could influence the outcome.

My second point is almost a converse of the first. While markets generate information through bargaining, this process is certainly not costless. Someone has to invest effort into finding the information required to reach an agreement of the kind described in Equation (7). Also, the bargaining process may itself generate undesirable rent seeking as the parties try to influence the outside options in order to get a bigger share of the total (Milgrom and Roberts, 1992). Such information and bargaining costs need to be compared with alternative ways of generating performance information. It is likely that in many situations, particularly ones with few participants, information can be produced in less costly ways outside the market. To go back to Alchian and Demsetz' original idea, perhaps the owner-monitor has expertise in evaluating the contributions of the input suppliers, making internal monitoring more efficient than allowing the parties to bargain privately. In this respect the assumption that the parties can observe ex post either the investment levels made or the actual value of the outside options deserves more scrutiny.

3.4 Empirical Issues[11]

The prediction of the Hart–Moore model that joint production never is optimal is counterfactual. Joint ventures (and shared ownership, more generally) have always been an important part of the corporate landscape, and if anything they seem to have become more prevalent in recent years [for instance, Stuckey (1983) reports that more than 50% of the world's primary aluminum was produced through joint ventures in the 1980s]. In fact, the theoretical predictions concerning joint ventures, sole ownership of complementary assets and outside ownership are all quite fragile. Multidimensional investments, as discussed later, can overturn these results. Different bargaining rules and dynamic considerations can alter the conclusions (Halonen, 1994; Baker, Gibbons, and Murphy; 1997, DeMeza and Lockwood, 1998). Also, if one assumes that investments add value to alienable assets rather than just human assets, things can change. Consider two parties that invest in a single alienable asset. If only one of them owns the asset, the other party will have no incentive to invest. If the investments are sufficiently complementary, a better arrangement is to let both sides veto the use of the asset. This will give each side a marginal return of one-half of v and both will invest. Of interest, joint ventures with a 50:50 ownership are quite common.

Sensitivity to details about bargaining rules and about the nature of the unobserved investments is likely to be empirically problematic. Another potential problem for property rights models is the interpretation of asset specificity. Consider again the competitive case where $v = v_b + v_s$ for all investment deci-

11. Whinston (1997) offers a much more detailed discussion of the empirical implications and problems with the property rights approach. Holmstrom and Roberts (1998) describes recent trends in organizational forms in reference to property rights and transaction cost theories.

sions. There is no asset specificity and, as we saw earlier, investment levels will be efficient. However, the same is true if one adds an arbitrary constant to the joint surplus, while leaving the outside options the same. In other words, there can be arbitrary amounts of asset specificity (measured by the constant) without any effect on the optimal distribution of assets. Again, this is an implication that is sensitive to the particular bargaining game being used. Even so, it is clear that the meaning of asset specificity in the Hart–Moore model, if there is one, is very different from the standard definition underlying all the empirical work on transaction cost economics. It cannot be measured by ex post dependency as in Williamson (1975, 1985). Rather, the degree to which asset specificity is a problem has to do with how investments influence the marginal values v, v_b, and v_s. It is challenging to think of ways to measure these effects, since they involve investments and associated margins that according to the theory cannot be observed or priced.

3.5 Why Do Firms Own Assets?

Let me turn to the most serious question concerning the model. To what extent is it about the firm?

The problem is that the theory, as presented, really is a theory about asset ownership by individuals rather than by firms, at least if one interprets it literally. Assets are like bargaining chips in an entirely autocratic market. There are neither firms nor workers. The same critique that was directed at Alchian and Demsetz' vision of the firm, that organizational affiliations did not matter for transactions, could be directed at the Hart–Moore model just as well. Individual ownership of assets does not offer a theory of organizational identities unless one associates individuals with firms. So why not think of the model as an entrepreneurial theory of the firm? Suppose unobserved human capital investments are investments in some sort of organizational capital or capabilities. The main problem with this interpretation is that it takes firms and their activities as exogenously given. If one tries to explain how assets get distributed across firms, given a set of investment activities, this is all right. But if one tries to explain how activities get distributed across firms, it obviously is not. And yet the boundary question is in my view fundamentally about the distribution of activities: What do firms do rather than what do they own?[12] Understanding asset configurations should not become an end in itself, but rather a means toward understanding activity configurations.

Going back to the literal (individualistic) interpretation of the model, its logic would seem to suggest that assets should be distributed widely across individuals, just as shares in the joint output would typically be widely distributed in the free-rider model. If we look inside firms, we should see workers own assets either outright or perhaps intermittently in some way in order to create optimal incentives. Clearly we don't see anything like that. Instead, the overwhelming

12. This agrees with Coase's (1988) view of what needs explaining. Levin (1998) offers a model in which activities get clustered because of externalities in reputation building.

evidence is that firms own essentially all the productive alienable assets and employees rarely any.

One can present this critique a bit more pointedly as follows. When firms own assets, these assets are tied together in a fashion that makes it impossible for a person to own 30% of one of the firm's assets and 50% of another of the firm's assets. If I own 10% of firm A, I own 10% of all its assets. But in a model where assets provide the only incentives, it is clear that this is generally inefficient. It would always be weakly better to have each asset be separately sold and available for individual purchase. Such an arrangement would permit maximum flexibility in distributing ownership shares in assets. It cannot be worse, since if indeed I want to own 10% of all the assets of firm A, I could purchase 10% of each asset in the free market.[13]

So why do firms own essentially all the nonhuman assets it uses in production? Why do workers—or for that matter any other stakeholder—rarely own any such assets? This strikes me as one of the most basic regularities that a theory of the firm needs to explain.

One possible explanation is that ownership strengthens the firm's bargaining power vis-à-vis outsiders. Suppliers and other outsiders will have to deal with the firm as a unit rather than as individual members. It is readily seen that in the Shapley value formula (which Hart and Moore use in the multiperson case), a coalition that supplies substitute inputs will get more as a collective bargaining unit than individually. United the firm's members stand stronger than divided, just as a union can extract better terms for its members. On the other hand, if the inputs are complements, then workers can extract more of the surplus (according to Shapley value) by bargaining individually, so this line of reasoning is not so clear-cut. The general point though is that institutional affiliation, and not just asset allocation, can significantly influence the nature of bargaining. Note that this type of argument does not require any ex ante investments. Ownership coalitions could reflect the equilibrium of a power game.

Another plausible and related reason for bundling assets is that it influences the terms of financing asset purchases. Integration can assure the financier that there are fewer holdups if the financing terms ever have to be renegotiated. Also, with assets that can collectively be pledged as collateral, the firm may increase its financial leverage and this way obtain more as well as cheaper financing (Conning, 1997). In effect, the firm can be a valuable financial intermediary. The recent literature on corporate finance is relevant here, but there have been very few models that try to characterize the division of labor between direct market finance and firm-mediated finance.

A third reason is that by having the assets under a single authority (at least formally) the firm can assign workers to assets in a manner that is richer and more varied than what would be possible with separate ownership. The firm

13. This statement may need qualification, if as seems likely, there are externalities in the market for assets. See Bolton and Whinston (1993).

can shift workers around over time and across contingencies because ownership of assets gives it the right to design jobs associated with those assets. The firm can promote workers it deems deserving and it can demote or fire workers it thinks have underperformed, but only to jobs that are associated with assets that it owns. A firm cannot promote a worker to a job in another firm, nor can it typically assign a worker to a job that requires assets controlled by another firm. The fact that contingencies are hard to specify ex ante makes the firm a potentially important operator of an internal human capital market.

This last point is part of a broader explanation. By focusing on holdups alone, the property rights approach overlooks the great variety of instruments that can be used to influence employee incentives. If all a firm can do is pay its employees based on their holdup power, its ability to provide incentives will be quite limited. Actually the firm has a great variety of instruments with which to influence worker incentives. The firm can pay bonuses, based on either an explicit formula or subjective performance evaluations. It can design jobs, set work rules, assign tasks, decide on strategic plans, delegate authority, control information channels, espouse a particular corporate culture, and so on.[14] What gives the firm the power to dictate all these things? The reason is apparent: by owning and hence holding contracting rights to all the alienable assets, the firm gains considerable leverage over its human assets. Hart and Moore suggest likewise, but in their case the power is merely used to alter the constellation of holdups and the resulting bargaining payoffs rather than to give the firm access to a wider range of instruments. Indeed, their logic fits better the first explanation given above, that concentrated ownership gives leverage over outsiders.

Why should the firm have any such power? My argument is that it allows the firm to internalize many of the externalities that are associated with incentive design in a world characterized by informational imperfections. As the theory of second best suggests, an uncoordinated application of the available incentive instruments will lead to significant externalities. By having access to more instruments, the firm can set up a more coherent system of incentives. Often this involves suppressing excessively strong incentives on individually measured performance for the benefit of enhancing the effectiveness of more delicate and subtle instruments aimed at encouraging cooperation and other less easily measured activities. In other words, high-powered ownership incentives are replaced by lower-powered employee incentives, because that enables a better overall design. This view stands in sharp contrast with the commonly heard complaints about lack of individual initiative in large hierarchical firms. That type of critique does not appreciate the inevitable trade-off between individual initiative and cooperation when performance is imperfectly measured.

Let me stress that viewing the firm as a subeconomy, which regulates trade according to second-best principles, does not imply that one firm should own

14. Of course, a firm could and does pay bonuses to independent contractors as well (and implicitly "promote" such contractors by awarding them increasingly valuable jobs), but the evidence suggests that the nature of external contracting is quite different from internal contracting.

all the assets. As the property rights approach shows, separate ownership does allow market-based bargaining, which cannot be emulated within a firm that owns all the assets. Indeed, the very fact that workers can exit a firm at will and go to other firms, and that consumers and input suppliers and other trading partners can do likewise, limits the firm's ability to exploit these constituents. This in turn makes it feasible for the firm to go about its business of setting "internal rules of the game" in a relatively unfettered fashion.[15] By comparison, a government, which faces a similar regulation problem, does not have the luxury to use nonanonymous rules. Its regulations must apply equally to all firms and citizens that face verifiably similar circumstances. Also, voice has to replace exit as the instrument that limits government abuse (Hirschman, 1970). Democracy is a costly governance procedure, but it appears to be the best one available when exit is precluded.[16]

4. Regulating Trade Within the Firm

4.1 Various Measurement Costs

In this section I will illustrate some of the principles and implications of coherent incentive design based on Holmstrom and Milgrom (1991, 1994). I will use a moral hazard model in which explicit incentive payments are possible, but it will become evident that the basic insights I am after do not depend on this feature. In particular, other incentive drivers, such as asset ownership, can often replace explicit contracts without altering the main point.

Moral hazard stems from imperfect performance measurement. There are (at least) three types of measurement imperfections one can envision. The standard one is that performance is a joint product of what the agent does and what nature does:

$$x = f(e, \theta),$$

where e is the agent's choice of input, θ is the realized state of nature, and x is the observed performance (which may not equal output y). Now the problem is akin to joint production in the free-rider model: one doesn't know exactly how to apportion blame (or praise). This is not a problem if the agent is risk neutral and has sufficient wealth to buy out the productive assets and become a residual claimant. But if the agent is risk averse or has too little wealth, there is a trade-off between stronger incentives and sharing of risk.

The second type of measurement problem arises when performance measures are biased. Suppose

$$x = \theta e$$

15. Thus markets and firms are not mere substitutes, as Coase argued, but also complements insofar as strong market discipline permits the successful operation of large and powerful firms. This is an aspect that the organizational literature has completely overlooked, because the focus has been on individual transactions rather than on the system as a whole.

16. Large firms feel similar pressures to be democratic and nonpartial when exit is costly for its members.

and output $y = e$. Assume that the agent can observe θ before he chooses e, that the unobserved cost function $c(e)$ is strictly convex, and that in expectation x is an unbiased measure of output: $E(\theta) = 1$. The problem then is that the agent's choice will vary with θ even though the marginal product of his choice in the production function is constant. Even if the agent were risk neutral, paying him according to the sharing rule, $s(x) = x+$ constant, would be inefficient because this rule induces excessive variation in effort choice, which is costly due to the convexity of the cost function c (Baker, 1992). The situation is very similar to the one analyzed many years back by Weitzman (1974) in his seminal article on prices and quantities [for more on the comparison, see Courtey (1997)]. There is a misalignment between what is measured and what is produced.

The third type of measurement cost stems from the agent's ability to manipulate the performance measure (Holmstrom and Milgrom, 1991; Allen and Gale, 1992). Let

$$x = e + m, \tag{9}$$

where m is the (unobserved) degree to which the agent manipulates x. If manipulation is privately costly, then this case will also lead to contracting costs even though the agent is risk neutral. I will use this framework, because it is convenient and not unrealistic. One interpretation of "manipulation" a la Equation (9) is that the agent can shade on quality: real quality is e, while observed quality is x, with m measuring the degree of shading.

4.2 A Simple Model

To be concrete, assume that the costs of effort and manipulation are separable and take the form

$$\text{private cost} = c(e) + d(m) = 1/2(e^2 + \lambda m^2), \tag{10}$$

where λ is a parameter measuring the relative cost of manipulation. Let the value of output be $y = pe$, where p is the value of the agent's input. I restrict attention to linear sharing rules of the form $s(x) = \alpha x + \beta$. In this setting of certainty there is no loss of generality in doing so. (Quantity constraints would do just as well; the incentive on x would be represented by a shadow price.) The agent's response is to choose $e(\alpha) = \alpha$ and $m(\alpha) = \alpha/\lambda$. Note that the ratio $e/m = \lambda$ will be the same no matter what incentive scheme is used (even if it is nonlinear).

The optimal choice of α is found by maximizing total surplus

$$S(\alpha) = p(e(\alpha)) - c(e(\alpha)) - d(m(\alpha)) = p\alpha - 1/2(\alpha^2 + \alpha^2/\lambda). \tag{11}$$

The best choice is to set $\alpha = \lambda p/(1 + \lambda)$. This contrasts with the first-best choice, which is $\alpha = p$. In the second-best choice, the incentive provision is always weaker than in first-best.[17]

17. By using total surplus as my welfare measure, it would seem that I have made the principal carry part of the agent's direct cost of cheating. Can't the principal just give the agent an incentive scheme with the surplus splitting parameter β set so that the agent meets his participation constraint

This model isn't very interesting as such. But it provides the seed for discussing richer variations, namely variations in which the agent's input e is multidimensional, as when the agent can engage in many different tasks. In reality, the problem of providing sufficient effort may often be of second-order importance to the problem of having the agent allocate effort efficiently among several tasks. For simplicity I will introduce multitasking with just two measured tasks and assume that they represent time spent on each task. Denote the two inputs e_1 and e_2 and let the new cost function be

$$\text{private cost} = 1/2(e_1 + e_2)^2 + 1/2\lambda m^2. \tag{12}$$

Let the principal's return function be $y = p_1 R(e_1) + p_2 e_2$, where R is an increasing, strictly concave function (in order to avoid corner solutions). The inputs can be observed via two performance measures:

$$x_1 = R(e_1)$$
$$x_2 = e_2 + m.$$

I have simplified things by not letting the payoff from e_1 be manipulable; the effects of adding manipulation to x_1 will be evident. The return from e_2 can be manipulated at the marginal cost λm.

Payments to the agent take the form $s(x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2 + \beta$. The choice of the optimal incentive coefficients proceeds as before. The first step is to determine how the agent responds to an arbitrary pair of incentive coefficients (α_1,α_2). The agent chooses e_1, e_2, and m to

$$\max \alpha_1 R(e_1) + \alpha_2(e_2 + m) - 1/2(e_1 + e_2)^2 - 1/2\lambda m^2.$$

The first-order conditions for e_1 and e_2 are:

$$\alpha_1 R'(e_1) = \alpha_2,$$
$$e_2 = \alpha_2 - e_1. \tag{13}$$

The first-order condition for m is

$$m = \alpha_2/\lambda. \tag{14}$$

Of course, the principal can figure out exactly the amount of manipulation that takes place, so he is not fooled. Yet there is a deadweight loss from the manipulation activity, just as in any signaling model. The loss is

$$1/2\lambda m^2 = \frac{\alpha_2^2}{2\lambda}, \tag{15}$$

which increases as λ is reduced. The easier it is to manipulate the performance measure, the more wasteful this activity becomes.

The second step in the analysis is to choose the incentive coefficients α_1 and α_2, given the agent's response functions [Equations (13) and (14)]. There is little point in showing the details, so let me just report the main features. In the

only if he does not manipulate? Not so. If such a scheme is chosen, the agent will actually do strictly better than his reservation utility.

first-best solution (corresponding to infinitely costly manipulation; $\lambda = \infty$), the principal sets $\alpha_1 = p_1$ and $\alpha_2 = p_2$. In the second-best solution, incentives on e_2 are muted (low-powered) in order to reduce wasteful manipulation. This makes it optimal to set weaker incentives for e_1, as well, because e_1 and e_2 compete for the agent's attention. It is easy to check that both activities will in fact be performed at lower levels of intensity than in the first-best solution. Also, as manipulation becomes less costly to the agent (λ decreases), the distortion in inputs will increase, making it desirable to reduce the incentives further; this follows from Equation (15), using a revealed preference argument. The economic point is that when performance measures have less integrity, there will be more wasteful manipulation. To reduce the waste without creating a significant imbalance in the allocation of effort, it is optimal to lower incentives on both activities.

An instructive way to think of this second-best result is that there are two ways to increase incentives for the problematic task e_2. Either the principal can increase the reward on e_2 or he can reduce the reward on e_1. The benefit of the latter alternative is that, by reducing the opportunity cost of e_2, incentives improve without any increase in manipulation. The general idea that an optimal incentive design should consider not only rewards but also instruments for influencing the agent's opportunity cost is evident in the many constraints firms impose on their employees.

4.3 Freedom and Constraints

A more direct way of seeing the role of the opportunity cost is by considering the following variant. Go back to the one-task case, but add a parameter τ in the agent's cost function so that now

$$\text{private cost} = c(e; \tau) + d(m; \lambda).$$

The agent will choose e and m so that

$$c'(e; \tau) = \alpha = d'(m; \lambda).$$

Thus as α goes up, both e and m go up. For simplicity, let d be of the same functional form as in Equation (12). For c, assume that the second derivative with respect to e is constant and denote it $C''(\tau)$. Since $de/d\alpha = 1/C''$, C'' measures how responsive the agent is to the incentive coefficient α. Assume further that $dC''/d\tau < 0$, so that a higher τ corresponds to a more responsive agent, and that the marginal cost c' increases with τ. These assumptions will be satisfied, for instance, if the agent can engage in (unobserved) private activities as well as in the principal's activity e, and only the total time spent on all activities determines the agent's cost of effort (that is, total time is limited by a set number of hours). In that case, if τ represents the number of private tasks that are permissible, the conditions stated above will be met. In this case, we can think of τ as measuring the agent's freedom from rules and restrictions. Other interpretations are possible. For instance, τ could represent the choice of a less or more flexible technology.

Total surplus, expressed as a function of the endogenous design variables α

and τ and the exogenous parameters, p and λ, is

$$TS(\alpha, \tau; p, \lambda) = pe(\alpha, \tau) - c(e(\alpha, \tau); \tau) - d(m(\alpha, \lambda); \lambda). \tag{16}$$

In order to see how the optimal design changes with changes in the parameters, one must consider the interaction effects of the variables in TS. To illustrate, consider the interaction terms with α. We have

$$dTS/d\alpha = (p - \alpha)\partial e/\partial\alpha - \alpha\partial m/\partial\alpha, \tag{17}$$

which is set to zero at the optimal value of α. Taking the derivative of Equation (17) with respect to τ, we see that $\partial^2 TS/\partial\alpha\partial\tau > 0$, because $\partial^2 e/\partial\alpha\partial\tau$ is positive (by assumption), the elasticity of effort is constant, and $p > \alpha$ (in any meaningful design). For the other cross-partials we find that $\partial^2 TS/\partial\alpha\partial p > 0$ and $\partial^2 TS/\partial\alpha\partial\lambda = 0$. Also, $\partial^2 TS/\partial\tau\partial p < 0$ and $\partial^2 TS/\partial\tau\partial\lambda = 0$.

As we can see, the total surplus function is not supermodular in all variables. This is not a problem in predicting the effects of changes in λ, the cost of manipulating performance. It will unambiguously increase α as well as τ. The less there is of a measurement problem, the stronger the incentives and the more freedom the agent will be given. Now consider changes in p, the value of the agent's effort. Here we encounter a problem. The problem is that an increase in p calls for a strengthening of incentives through a higher α, and a reduction in the opportunity cost through a lower τ, yet when α increases τ should also increase, other things equal. The two effects go in opposite directions and the prediction becomes ambiguous. If the interaction between α and τ is strong enough, an increase in p will raise both α and τ. If the effect is sufficiently weak, an increase in p will raise α and lower τ.

These comparative exercises are meant to illustrate two points. The first is that with more than one instrument, one needs to think about coherence in the overall design. Contracting separately on the different instruments will create externalities. The second, more subtle point is that predictions regarding coherence can be quite sensitive to the external factors that drive the change. Thus if the principal's activity becomes more important, it will imply stronger emphasis on performance, but whether that is followed by more or less freedom depends on how strong the interaction between α and τ is.

Firms extensively use the option to constrain their employees. Many of these constraints are so obvious as to elude our recognition. For instance, it is standard praxis that ordinary employees cannot work for two firms at the same time. Also, while working at home may be becoming more common, the majority of employees are still required to come into the office during work hours. Another example is the restriction on the use of money as a means to curry favors and compete for services within the firm.[18] Allowing employees to do so would diminish or destroy some of the more delicate incentives employed by firms, such as social norms.

18. Excessively generous gifts to a shared secretary are well known to cause tensions, for instance. In a related vein, Prendergast and Stole (1997) discuss why money can be harmful in maintaining effective social and reputational networks.

5. Back to Boundaries

The previous section dealt with designing incentives within a single firm. It made the point that incentive design can benefit greatly from controlling a wider range of instruments. Because many of the instruments are accessed through ownership of assets, this perspective has implications for the boundary of the firm. If the instruments are under the control of separate firms, it is more difficult to have them designed in a coordinated fashion than when the instruments are controlled by one firm. Separate design will create contractual externalities, which integration can help to resolve. Let me illustrate this idea with two variations on the first of the examples from the previous section.

5.1 Variation 1—Asset Ownership[19]

In this variation incentives stem exclusively from asset ownership, as in the Hart–Moore model. There are two agents, 1 and 2. There is one asset A. The asset can only be operated by agent 1, who can produce outputs $y_1 = R(e_1)$ and $y_2 = e_2$ with it, choosing unobserved inputs (e_1,e_2). The agent's private cost function is given by Equation (12) with $m = 0$, so manipulation is not possible. By bargaining with an outsider, agent 1 can get p_1 for each unit of output y_1. Agent 2 is an expert on marketing output y_2, but has no investment decisions. His human capital is essential for selling output y_2 at a positive price p_2 per unit.

In keeping with incomplete contract models, assume that revenues cannot be verified and hence contracts based on y_1 and y_2 cannot be written. Instead, revenues from y_1 and y_2 get divided between the agents through a bargaining process. The division depends on who owns the only alienable asset A. If agent 1 owns asset A, he can appropriate $1/2 p_2 y_2$ from agent 2's revenue by threatening to withhold output y_2. In addition, he will get $p_1 R(e_1)$ from his bargain with outsiders, since production of y_1 does not involve agent 2. Agent 2's payoff is $1/2 p_2 y_2$. If agent 2 owns asset A, the payoffs to agent 1 are determined by his right to withhold his human capital in the production of y_1 and y_2. With a 50-50 split of the surplus, he will appropriate $1/2[p_1 R(e_1) + p_2 e_2]$. The same amount will go to agent 2.

With these reduced form payoffs, the analysis maps into the general framework laid out in Section 4.3. If agent 1 owns asset A he faces incentive coefficients $\alpha_1 = p_1$ and $\alpha_2 = 1/2 p_2$. If he doesn't own the asset, the implied incentives are $\alpha_1 = 1/2 p_1$ and $\alpha_2 = 1/2 p_2$. Agent 2 makes no investments so the optimal ownership structure is determined entirely by the effect it has on agent 1's incentives. Agent 1's allocation of effort when he owns A is given by

$$\begin{aligned} e_1 &= p_1 R'(e_1) - e_2, \\ e_2 &= \tfrac{1}{2} p_2 - e_1. \end{aligned} \tag{18}$$

It follows that e_1 is above the efficient level of effort, while e_2 is below it. When

19. This variation was inspired by Rajan and Zingales (1998).

agent 1 does not own A, the allocation of effort is

$$\begin{aligned} e_1 &= \tfrac{1}{2} p_1 R'(e_1) - e_2 \\ e_2 &= \tfrac{1}{2} p_2 - e_1. \end{aligned} \tag{19}$$

Now e_1 is at the first-best level, while e_2 is below it. However, e_2 is closer to first-best in Equation (19) than in Equation (18). Consequently, the case where agent 1 *does not* own asset A provides strictly better incentives than the case where he does own A, irrespectively of parameter values. The logic is the following. When agent 1 does not own the asset, incentives have the right balance. Since the choice of e_1 only depends on the right balance, it will be chosen efficiently. While e_2 will be below the first-best level, it will be higher than when agent 1 owns the asset and α_1 equals p_1, because in the latter case e_1 is too high, driving down e_2.

The purpose of this example is to illustrate—with asset ownership rather than explicit payments as the incentive driver—that balanced incentives, even when low powered, can be better than imbalanced incentives. Ownership by agent 1 can be costly, not only because the asset cannot then be used to motivate agent 2; indeed, agent 2 needs no motivation here. Rather, ownership can be directly detrimental for agent 1's incentives because it deflects attention from activity e_2 . For this reason, it may be desirable to have assets owned by a third party with no inputs at all (agent 2 can act as a third party here). It is easy to envision an extension of the example in which assets get clustered in the hands of a third party in order to remove excessively strong incentives that deflect attention to outside opportunities (Rajan and Zingales, 1998). The third party could be a firm, but there is no particular reason for that. Any individual would do as well.[20] In either case, the logic of integration is not one of asset complementarities as defined in Hart and Moore (1990), but rather of incentive complementarities caused by contractual externalities as in Holmstrom and Milgrom (1994).

A final note on this case: The result that agent 1 should never own asset A depends on the assumption that surplus is split 50-50 in the bargain between agents. Suppose agent 1's human capital is not indispensable, so that agent 2, if he owns the asset, could find a substitute for agent 1. As long as agent 1 is better than the substitute, he will be retained and get a share of the y_2 revenue, though less than 1/2 of it. The better the substitute is, the smaller agent 1's share of the surplus will be. Eventually there will come a point when it is better to have agent 1 focus solely on maximizing y_1. At that point, agent 1 should own asset A.

5.2 Variation 2—Contracts and Job Design

Let me return to the example in Section 4.2, reinterpreting it in an institutional context. A worker works for firm A. He can allocate his time between two

20. This is a form of beneficial budget-breaking in the context of ownership.

activities e_1 and e_2. Both tasks require an asset that firm A owns and therefore only firm A can contract with the worker for the supply of e_1 and e_2. Outputs are, as before, $y_1 = R(e_1)$ and $y_2 = e_2$. Firm A can sell output y_1 to an outside market at price p_1 per unit. Output y_2 is used exclusively by firm B, which values it at p_2 per unit. Contracting between firm A and the worker can be based on the performance measures $x_1 = y_1$ and $x_2 = y_2 + m$, where m is a variable that the worker can choose to manipulate actual output. Assume that by virtue of its contracting rights, only firm A can observe the contract it makes with the worker and that firm B cannot offer the worker any secret side payments. This is in accordance with employment laws. I will shortly suggest why such side payments may indeed be undesirable. The only way firm B can induce the worker to pay attention to e_2 is by signing a contract with firm A based on the observed measure x_2. I do not allow firm B to pay based on x_1. In practice, firm B is unlikely to observe firm A's internal performance measures. More importantly, firm B cannot control the worker's assignment, so if it tried to discourage the worker from working on e_1 through a contract on x_1 (in order to increase attention on e_2) firm A could reassign the worker to an alternative, third task, with the same effect as if firm B could not use x_1; (or firm A could side contract with the worker to mitigate firm B's influence).

The analysis now proceeds as in Section 4.2 with the difference that firm B only has access to one incentive variable: the piece rate contract $rx_2 + b$ that it uses to reimburse firm A for the services of the worker. Taking this piece rate contract as given, firm A designs an incentive scheme (as in Section 4.2) to induce the worker to choose effort and manipulation in a way that maximizes the joint surplus $p_1R(e_1) + r(e_2 + m) - 1/2\lambda m^2 - 1/2(e_1 + e_2)^2$. This surplus is the sum of the worker's and firm A's payoffs. The worker's incentives are determined by Equations (13) and (14) as before. What coefficients α_1 and α_2 will maximize the joint surplus? Evidently, the optimal choice is $\alpha_1 = p_1$ and $\alpha_2 = r$, since then Equations (13) and (14) will coincide with the first-order conditions for maximizing the joint surplus by direct choice of e_1, e_2, and m.

Return now to firm B's choice of a piece rate contract. Suppose manipulation is infinitely costly. Then the first-best outcome will obtain by setting $r = p_2$. But as manipulation becomes more attractive (λ decreases), there will be more waste and firm B will decrease r in response. (Note that firm B cannot decrease r at the same rate as if it controlled both incentive coefficients, because e_2 has to compete with e_1, and firm A keeps the incentive for e_1 at p_1.) Eventually, as the misallocation of effort gets sufficiently large due to a mispricing of the returns to effort, firm B will stop buying y_2 from firm A and choose one of two alternatives. It will either buy the asset from firm A in order to get access to both incentive instruments α_1 and α_2 and that way get a better handle on the effort distortion problem. Or it can simply give up on y_2 altogether. Which of the two alternatives firm B chooses depends on the relative values of y_1 and y_2. Actually, if firm B chooses to buy the asset, it also has the option to start producing y_1 for firm A in a reversal of the originally envisioned trade. Indeed, if firm B could appropriate all of p_1y_1 in such a trade, it would be optimal (first-

best, in fact) to have firm B own the productive asset and employ the worker, since y_1 can be measured without error and hence there are no contractual externalities. However, assume that firm A is run by an entrepreneur whose services are essential for getting the return from y_1. In that case firm B will appropriate less than the social surplus by bargaining with the entrepreneur and this again creates distortions, possibly of such magnitude that it will not be worthwhile for firm B to produce y_1 at all.

In sum, depending on the degree of measurement costs and the returns to the various activities, we can have three different configurations: (a) the worker could be working for firm A, which owns the asset and produces both y_1 and y_2; (b) the worker could be working for firm B, which owns the asset and produces both y_1 and y_2; or (c) either y_1 or y_2 are not produced at all.

The point of this example is to show that measurement costs can play an important role in determining boundaries. A way to express the measurement problem illustrated by the example is to say that market instruments, such as output prices, do not always allow agents to communicate their needs and desires sufficiently well. With measurement costs of the kind described here, a price contract does not permit firm B to express and get exactly what it wants from firm A. Ownership of the asset expands on firm B's ability to communicate its preferences to and extend its leverage over the worker. This contrasts with Bernheim and Whinston's (1986) work on common agency. [My example is also one of common agency in that both firms (principals) have an interest in influencing the worker (agent).] In Bernheim and Whinston's model, the problem is resolved by having all principals transfer their preferences to a single principal along with the decision rights. This internalizes the common agency externality because the contracts between the principals incorporate all means of influence.

Of course, markets use more instruments than prices to influence incentives and mediate transactions. It is easy to come up with examples where market transactions specify exclusion of activities, choice of technology, special work assignments, and so on. For instance, firms may insist on being served by a particular consultant in a larger agency; car manufactures, including General Motors, typically insist on expensive quality measures, such as the installation of an ISO 9000 standard, before certifying a supplier; and recently, in a break with tradition, Sears has started to demand that incentives paid to their suppliers be passed onto those individually responsible for the work. These kinds of measures are evidently taken to assure quality through inputs rather than outputs. Does that nullify the ideas expressed above? Not really. First, these examples show that information problems of the sort I have emphasized do exist. Second, the theory does not assume that only prices can be used in markets, even though the examples were designed that way. It merely instructs us to look at the relative costs of controlling incentives within firms versus on markets, suggesting that control of assets provides control over a number of indirect incentive measures that typically, but not always, can be inaccessible or very costly to use in the market.

Finally, note that in this illustrative example, the boundary is most naturally

defined by the set of activities undertaken by each firm. In a simple extension, one can see that if, for instance, firm B only ends up undertaking activity y_2, firm A may want to buy a duplicate asset and hire its own worker to produce y_1, underlining the sense in which asset ownership follows the choice of activities rather than the reverse. In most cases, if one were only told what assets a firm owns, but not what it is doing, it would be difficult to guess its lines of business. But if one were told what line of business the firm is engaged in, it would in many cases be easier to suggest the types of assets it owns.

5.3 Measurement Cost versus Holdup—Empirical Distinctions

The theory advocated here argues that boundaries will be drawn with intent to internalize significant contractual externalities stemming from measurement costs. How does this differ empirically from the predictions of the holdup logic? Before answering that question, note first that the holdup problem is also one that involves externalities. The investments cannot be observed and the future contingencies cannot be contractually specified. Integration solves the problem by allowing the firm to structure the incentives for its managers so that they come to terms with each other more easily than they would if they were in separate firms. This requires that the firm puts the managers inside the firm on weaker individual incentives in order to improve cooperation or, alternatively, that it does not allow the managers any power to negotiate internal trades (see Holmstrom and Tirole, 1991). Both solutions are associated with costs, as individual incentives are muted.

Now, back to the comparison with holdups. Consider the following two scenarios. In one case, ex post trade is perfectly enforceable, but it cannot be specified ex ante. There are no measurement costs ex post, but there is a potential for holdup as there is just one supplying firm. In the other case, there are a large number of suppliers and potential buyers and hence no asset specificity. But each supplier-buyer relationship faces the sort of measurement problem described earlier. The holdup theory would predict integration in the first scenario, but not the second. Measurement cost theory would predict just the reverse. This thought experiment illustrates how the two ideas are empirically distinct, though they are by no means mutually exclusive.

One case in which the measurement story has been shown to be empirically relevant is the study by Anderson and Schmittlein (1984). They found that firms, when deciding between an in-house sales office versus use of an external sales firm, the key factor explaining this choice was the difficulties with measuring a salesperson's performance. The biggest problems occurred when the salesperson worked in teams assigned to sell many different brands. This is very much in the spirit of the examples above.[21]

21. Additional evidence can be found in Brickley (1995), Slade and Lafontaine (1995), and Slade (1996), among others. Evidence against measurement costs is reported in Wernerfelt (1997).

6. Concluding Remarks

The modern property rights theory, developed by Hart and his coauthors, has had a major influence on the evolution of the theory of the firm. By emphasizing incomplete contracting and residual decision rights, in place of imperfect contracts and agency costs, it has brought much needed clarity to the meaning of ownership and the significance of exit as a defining characteristic of the market. Thanks to this theory, we have an improved understanding of how market incentives operate and, importantly, a modeling technology with which to investigate them.

Despite these significant contributions, I have argued that the property rights theory, as articulated in Hart and Moore (1990) and other representative pieces, says very little about the firm.[22] The problem is that there really are no firms in these models, just representative entrepreneurs. This focuses all the attention on the question of boundaries, which admittedly has been the outstanding puzzle to solve ever since Coase raised it more than 50 years ago. However, we also have to ask why Coase thought (or we think) that the question of boundaries is such an interesting and essential one to address. In my view, the boundary question is interesting primarily because we cannot claim to fully understand either the internal organization of firms or the operation of markets by studying the two in isolation. We need to analyze how they interact as organizations; how they compete as well as complement each other in matching individuals with tasks and in providing proper individual incentives for carrying out those tasks. It is in this spirit that I have suggested that we take a broader view of ownership than is common in the current property rights models.

References

Alchian, Armen, and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization," 62 *American Economic Review* 777–795.

Allen, Franklin, and Douglas Gale. 1992. "Measurement Distortion and Missing Contingencies in Optimal Contracts," 2 *Economic Theory* 1–26.

Andersen, Erin, and David Schmittlein. 1984. "Integration of the Sales Force: An Empirical Examination," 15 *Rand Journal of Economics* 385–395.

Arrow, Kenneth. 1974. *The Limits of Organization*. New York: W. W. Norton.

Baker, George. 1992. "Incentive Measures and Performance Measurement," 100 *Journal of Political Economy* 598–614.

Baker, George, Robert Gibbons, and Kevin J. Murphy. "Relational Contracts and the Theory of the Firm," draft, MIT.

Bernheim, Douglas, and Michael Whinston. 1986. "Common Agency," 54 *Econometrica* 911–930.

Bolton, Patrick, and Michael Whinston. 1993. "Incomplete Contracts, Vertical Integration and Supply Assurance," 60 *Review of Economic Studies* 121–148.

Brickley, James. 1995. "Incentive Conflicts and Contracting: Evidence from Franchising," FR 95-26, Rochester University.

Bull, Clive. 1987. "The Existence of Self-Enforcing Contracts," 102 *Quarterly Journal of Economics* 147–159.

Coase, Ronald. 1937. "The Nature of the Firm," 4 *Economica* 386–405.

22. The paper by Rajan and Zingales (1998) is an exception.

_____. 1988. "Industrial Organization: A Proposal for Research," in *The Firm, the Market and the Law*. Chicago: University of Chicago Press.

Conning, Jonathan. 1997. "Joint Liability, Peer Monitoring, and the Creation of Social Collateral," mimeo, Yale University.

Courty, Pascal. 1997. "Strategy Communication and Measurement Systems," mimeo, Universitat Pompeu Fabra.

DeMeza, David, and Ben Lockwood. 1998. "Does Asset Ownership Always Motivate Managers? Outside Options and the Property Rights Theory of the Firm," 113 *Quarterly Journal of Economics* 361–386.

Grossman, Sanford, and Oliver Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Lateral and Vertical Integration," 94 *Journal of Political Economy* 691–719.

Halonen, Maija. 1994. "Reputation and Allocation of Ownership," STICERD Theoretical Economics Discussion Paper, London School of Economics.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Oxford: Clarendon Press.

Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm," 98 *Journal of Political Economy* 1119–1158.

Hirschman, Albert O. 1970. *Exit, Voice and Loyalty*. Cambridge: Harvard University Press.

Holmström, Bengt. 1982. "Moral Hazard in Teams," 13 *Bell Journal of Economics* 324–340.

_____, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," 7 *Journal of Law, Economics, & Organization* 24–51.

_____. 1994. "The Firm as an Incentive System," 84 *American Economic Review* 972–991.

_____, and John Roberts. 1998. "Boundaries of the Firm Revisited," 12 *Journal of Economic Perspectives* 73–94.

_____, and Jean Tirole. 1991. "Transfer Pricing and Organizational Form," 7 *Journal of Law, Economics, & Organization* 201–228.

Klein, Benjamin, Robert Crawford, and Armen Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," 21 *Journal of Law and Economics* 297–326.

Laffont, Jean-Jacques, and Jean Tirole. 1991. "Provision of Quality and Power of Incentive Schemes in Regulated Industries," in J. Gabszewicz and A. Mas-Colell, eds., *Equilibrium Theory and Applications*. Cambridge: Cambridge University Press.

Levin, Jon. 1998. "Multilateral Contracting and Organizational Design," draft, MIT.

Maskin, Eric, and Jean Tirole. 1999. "Two Remarks on the Property Rights Literature," *Review of Economic Studies*, forthcoming.

Milgrom, Paul, and John Roberts. 1992. "Bargaining Costs, Influence Costs, and the Organization of Economic Activity," in J. Alt and K. Shepsle, eds., *Perspectives on Positive Political Economy*. Cambridge: Cambridge University Press.

Prendergast, Canice, and Lars Stole. 1997. "Monetizing Social Exchange," draft, University of Chicago.

Rajan, Raghuram, and Luigi Zingales. 1998. "Power in a Theory of the Firm," 113 *Quarterly Journal of Economics* 387–432.

Simon, Herbert. 1951. "A Formal Theory of the Employment Relationship," 19 *Econometrica* 293–305.

Slade, Margaret. 1996. "Multitask Agency and Contract Choice: An Empirical Exploration," 37 *International Economic Review* 465–486.

_____, and Francine Lafontaine. 1995. "Retail Contracting: Theory and Practice," draft, The University of British Columbia.

Stuckey, John. 1983. *Vertical Integration and Joint Ventures in the Aluminum Industry*. Cambridge: Harvard University Press.

Tirole, Jean. 1994. "Incomplete Contracts: Where Do We Stand?" Walras-Bowley lecture delivered at the North American Summer Meetings of the Econometric Society, Quebec City.

Weitzman, Martin. 1974. "Prices vs Quantities," *Review of Economic Studies* 477–491.

Wernerfelt, Birger. 1997. "On the Nature and Scope of the Firm: An Adjustment Cost Theory," 70 *Journal of Business* 489–514.

Whinston, Michael. 1997. "On the Transaction Cost Determinants of Vertical Integration," draft, Northwestern University.

Williamson, Oliver. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

———. 1979. "Transaction-Cost Economics: The Governance of Contractual Relations," 22 *Journal of Law and Economics* 233–271.

———. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.