Firm Scope v 32.wpd

"A Theory of Firm Scope"

by

Oliver Hart (Harvard University and NBER)

and

Bengt Holmstrom (MIT and NBER)

November 4, 2002*

*An earlier version of this paper circulated as "Vision and Firm Scope." The material presented here formed part of the first author's Munich Lectures (University of Munich, November 2001) and Arrow Lectures (Stanford University, May 2002). We are especially grateful to Andrei Shleifer for insightful comments. We would also like to thank Philippe Aghion, George Baker, Lucian Bebchuk, Pablo Casas-Arce, Mathias Dewatripont, Robert Gibbons, Meg Meyer, Chris Snyder, Jeremy Stein, Lars Stole, Eric van den Steen and seminar audiences at CESifo, University of Munich, Harvard University, London School of Economics, George Washington University, Stanford University, the Summer 2002 Economic Theory Workshop at Gerzensee, Switzerland and the University of Zurich for helpful discussions. Research support from the National Science Foundation is gratefully acknowledged.

Abstract

The existing literature on firms, based on incomplete contracts and property rights, emphasizes that the ownership of assets--and thereby firm boundaries-- is determined in such a way as to encourage relationship-specific investments by the appropriate parties. It is generally accepted that this approach applies to owner-managed firms better than large companies. In this paper we attempt to broaden the scope of the property rights approach by developing a simpler model with three key ingredients: (a) decisions are non-contractible, but transferable through ownership, (b) managers (and possibly workers) enjoy private benefits that are non-transferable, and (c) owners can divert a firm's profit. With these assumptions firm boundaries matter. Nonintegrated firms fail to account for the external effects that their decisions have on other firms. An integrated firm can internalize such externalities, but it does not put enough weight on the private benefits of managers and workers. We explore this trade-off first in a basic model that focuses on the difficulties companies face in cooperating through the market if benefits are unevenly distributed; therefore they may sometimes end up merging. We then extend the analysis to study industrial structure in a model with intermediate production. This analysis sheds light on industry consolidation in times of excess capacity.

1. Introduction

In the last fifteen years or so, a theoretical literature has developed that argues that the boundaries of firms--and allocation of asset ownership--can be understood in terms of incomplete contracts and property rights.[1] The basic idea behind the literature is that firm boundaries define the allocation of residual control rights and, in a world of incomplete contracts, these matter. In the standard property rights model, parties write contracts that are ex ante incomplete, but that can be completed ex post; the ability to exercise residual control rights improves the ex post bargaining position of an asset owner and thereby increases her incentive, and the incentive of those who enjoy significant gains from trade with her, to make relationship-specific investments; and as a consequence it is optimal to assign asset ownership to those who have the most important relationship-specific investments, or who have indispensable human capital.[2]

Although the property rights approach provides a clear explanation of the costs and benefits of integration, as a number of people have argued, the theory seems to describe owner-managed firms better than large companies.[3] There are several ways to see this. First, according to the theory, the major impact of a change in ownership is on those who gain or lose ownership rights; however, in a merger between two large companies it is often the case that the key decision-makers (the CEOs, for example) do not have substantial ownership rights before or after the merger. Second, the relationship-specific investments analyzed are made by individuals rather than by firms; this again resonates more with the case of small firms than large companies.

---

[1]See Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995). This literature builds on the earlier transaction cost literature of Williamson (1975, 1985) and Klein, Crawford and Alchian (1978).

[2]Extensions of the model show that it is sometimes optimal to take assets away from someone to improve their incentives to make relationship-specific investments (e.g., to discourage rent-seeking behavior). On this, see Baker, Gibbons and Murphy (2002a), Chiu (1998), de Meza and Lockwood (1998), and Rajan and Zingales (1998). For some recent empirical work supporting the property rights approach, see Baker and Hubbard (2002), Elfenbein and Lerner (2002) and Woodruff (2002).

[3]For a discussion of this and related points, see Holmstrom and Roberts (1998) and Holmstrom (1999).

Third, and perhaps most important, the approach envisions a situation of "autarchy," in which all the relevant parties meet and bargain ex post over the gains from trade, and the only issue is who has the right to walk away with which assets; there are no other decisions in the model. As it stands the model has no room for "organizational structure," "hierarchy" or "delegation"; in an important sense, the model continues to describe a pure market economy, although one enriched by the idea that individuals can be empowered through the ownership of key nonhuman assets.

The purpose of the current paper is to modify the property rights approach so that it can be applied to a broader set of organizational issues, including industrial organization and the organization of large firms. We begin by describing the key ingredients of our general approach.

Firms are characterized by the scope of their activities. We suppose that the economy has a given set of activities and we ask how these activities should be grouped to form firms. To be concrete, we will associate each activity with a production unit. We may think of a unit as an irreducible set of activities that it would be meaningless to break up further. Each unit has a manager and possibly also some workers. The owner of the unit is the boss. The unit manager could be the boss or the unit could be owned by an outside boss (sometimes referred to as a professional manager). The identity of the boss will be important.

The decisions in each unit are noncontractible (ex ante and ex post). Units can be bought and sold. Ownership of a unit confers the right to make the decisions in that unit. The fact that decisions can be transferred only through ownership, and are not even ex post contractible, is a central feature of our model and distinguishes it from the standard property rights model in which decisions are ex ante noncontractible but ex post contractible. Many decisions are of course contractible; for instance those involving trade between two firms. But there are probably many more decisions that firms make unilaterally, such as setting prices, choosing production techniques, deciding on marketing campaigns, assigning workers, determining incentive schemes, and so on. Some of these decisions have to be made unilaterally (e.g. pricing decisions); others are made unilaterally, because no one else has an interest in them. Our focus is on decisions that have potential external effects (positive or negative), yet are impossible to agree upon contractually. In these cases, changes in ownership are the only way to affect how decisions are made and how well they take into account external effects.

Full integration would always be optimal if all the benefits and costs from decisions were transferrable through ownership. To avoid this uninteresting conclusion, we assume that each unit generates two kinds of benefit: monetary profit and private (nontransferable) benefits in the form of job satisfaction for those working in the unit. For the most part we will assume that the manager is the only worker and hence private benefits refer to her job satisfaction. (We will return to discuss the source of private benefits in more detail later.) We assume that the boss of a unit can divert all the profit from that unit to herself. This simplifies the analysis by ruling out profit sharing as a way to influence incentives.

To illustrate how firm boundaries affect decision making, consider the simplest setting with two units A and B. Denote the pair of profits and private benefits (measured in money) accruing to each by (v_A, w_A) and (v_B, w_B), respectively. Then if the units are nonintegrated, and manager A is the boss of unit A, and manager B the boss of unit B, manager A will maximize $v_A + w_A$ since she diverts the profit from unit A and cares about her own private benefits; and manager B will maximize $v_B + w_B$ for similar reasons. In contrast, if units A and B are integrated, then what happens depends on who is the overall boss. If manager A is the overall boss, manager A will maximize $v_A + v_B + w_A$, since she diverts the profit from both units, and cares about her own private benefit but not manager B's. Similarly, if manager B is the overall boss, she will maximize $v_A + v_B + w_B$. Finally, if a (professional) outsider is brought in to be the boss, she will maximize $v_A + v_B$, since she diverts all the profit and does not care about private benefits.

As a reference point, note that social optimality is achieved by maximizing total surplus: $v_A + v_B + w_A + w_B$.

The important point here is that integration results in less weight being placed on private benefits than nonintegration. Under nonintegration w_A, w_B each appear in one boss's objective function. In contrast, under integration, at least one of the w's fails to appear in the overall objective function. However, this diminished influence of private benefits is offset by the fact that under integration <u>total</u> profits rather than individual unit profits are maximized.

In summary, under nonintegration, bosses maximize the right thing (profits plus private benefits) but are parochial (they do not take into account their effect on the other unit), while under integration they maximize the wrong thing but are broad.

We analyze two models of firm boundaries based on the general trade-off described above. The first, discussed in Section 2, is concerned with interactions between two units. To simplify matters, we suppose that each unit has a binary decision to make: it can either act in an accommodating way vis-a-vis the other unit or not. For example, the unit could accommodate by deciding not to poach on the other unit's territory or by being less aggressive in its advertising; or the unit could choose a technology or a standard that is beneficial to the other unit; or it could provide the other unit with some key piece of knowledge. Each of these examples has the characteristic that it may be too costly or impossible to agree on a decision contractually. Only by owning the unit can one control the decision.

We ask whether units A and B will be regulated better if the units remain independent or if they integrate. Call the case where both parties accommodate coordination. We show that, under reasonable assumptions, nonintegration leads to too little coordination; for some parameter values for which coordination is optimal, coordination will not be a Nash Equilibrium if units stay separate. There are two reasons why this may happen. The first is the well known prisoner's dilemma problem: both units may be better off with coordination, but unilaterally each prefers to deviate. The other reason is that while coordination may be socially optimal (total surplus maximizing), the distribution of gains may be so uneven that one of the parties will not want to go along (accommodate).

In contrast, under a weak assumption--specifically, that coordination represents a reduction in "independence" and therefore causes a fall in private benefits--integration leads to too much coordination. The reason is that an outsider boss may choose to coordinate even though it destroys significant private benefits, because it is more profitable.

In the model described above there is no role for outsourcing: decisions are always made inside the units and the only question is whether they are made by one boss or two. In reality, outsourcing of a decision or activity is often a viable alternative for an organization. For example, consider two units in the output market that need a specialized input. Then each unit can remain separate and provide the input in-house; or the units can merge horizontally and collectively provide the input in-house; or the units can remain separate and procure the input from an independent supplier.

In Section 3 we present a model that analyzes these choices (the model covers similar ground to Bolton and Whinston (1993), but the approach is different). We assume that the two units, or buyers, have stochastic input demands or valuations. The buyers can collectively choose whether to have one or two units of capacity in the upstream market. With two units of capacity, both buyers will be supplied, while with one unit only one buyer will be supplied. Given that valuations are stochastic, it may be efficient to save on capacity and have only one unit. However, saving on capacity typically requires coordination and specialization: the limited capacity must be dedicated or directed to the buyer with the greater needs. There are two leading organizational forms with one unit of capacity. In one the supplier is an independent firm ("outsourcing"); in the other the two buyers and a single seller merge into one firm ("complete integration"). (We also compare these with an organizational form with two units of capacity.) We show that in some cases potential savings in capacity can be exploited only under complete integration. The reason is that under other organizational forms the supplier may be unwilling to specialize to any one buyer. We argue that this model captures the idea of "integrating for synergy." Note that in order to allow for outsourcing we have to drop the assumption that ex post trade is noncontractible; instead we suppose that the decision to specialize to a buyer is noncontractible.

The interpretation that private benefits are enjoyed by a single manager is restrictive in that it implies that the units are in effect sole proprietorships under nonintegration. It is worth mentioning a second interpretation of the model that applies to the case where the units are large companies. One can interpret the private benefit w_i ($i = A,B$) as representing the aggregate job satisfaction of all the unit i workers rather than the private benefits of a single manager. Suppose that there is a population of bosses with different preferences, some of whom care about the firm's activities in addition to its profits, and others of whom care only about money. Call the first group <u>enthusiasts</u>, and suppose that there is an enthusiast corresponding to each unit: a unit A enthusiast has goals that are (partially) congruent with those of unit A workers, and hence puts some weight on w_A; while a unit B enthusiast has goals that are (partially) congruent with those of unit B workers, and hence puts some weight on w_B.[4] Assume as before that a boss can divert

[4]A boss may be biased toward its workforce because sustained contact with workers fosters friendship and empathy. Wrestling with the same problems, sharing the same information, and having a similar professional background are all conducive to a common vision that aligns interests, particularly on issues such as the strategic direction of the firm. Shleifer and

some fraction of the profit of any unit under her control to herself--but that this fraction is now less than 1. Then, under an appropriate assumption about the degree of congruence between bosses and workers and about the fraction, of profit diverted, a unit i enthusiast will maximize $v_i + w_i$ under nonintegration. In contrast, under integration, a professional will maximize $v_A + v_B$. We will show that this version of the model is analytically identical to the first one. Now the benefit of nonintegration for the owners of units A and B is that, since workers' private benefits are respected by an enthusiastic boss, workers will work for lower wages.[5]

Our paper is related to a number of ideas that have appeared elsewhere in the literature. First, there is an overlap with the recent literature on internal capital markets; see particularly Stein (1997, 2002), Scharfstein and Stein (2000), Rajan, Servaes and Zingales (2000), Brusco and Panunzi (2000), and Inderst and Laux (2000). This literature emphasizes the idea that the boss of a conglomerate firm, even if she is an empire builder, is interested in the overall profit of the conglomerate, rather than the profits of any particular division. As a result, the conglomerate boss will do a good job of allocating capital to the most profitable project ("winner-picking"). Our idea that the professional boss of an integrated firm maximizes total profit is similar to this; the main differences are that the internal capital markets literature does not stress the same cost of integration as we do--the insufficient emphasis on private benefits--or allow for the possibility that the allocation of capital can be done through the market (in our models, the market is always

Summers (1988) argue that it may be an efficient long-run strategy for a firm to bring up or train prospective bosses to be committed to workers and other stakeholders (on this, see also Blair and Stout (1999)). Milgrom and Roberts (1988) argue that frequent interaction gives workers the opportunity to articulate their views and influence the minds of their bosses, sometimes to the detriment of the firm. All these explanations are consistent with our assumption that the boss of a firm with broad scope will put less weight on private benefits than a boss of a firm with a narrow scope. The reason is that, with a broader range of activities, the firm's workforce will be more heterogeneous, making the boss experience less empathy for any given group. Also the intensity of contact with any particular group will go down, reducing the ability of that group's workers to influence the boss.

[5]There is some evidence consistent with this. Schoar (2001), in a study of the effects of corporate diversification on plant level productivity, finds that diversified firms have on average 7% more productive plants but pay their workers on average 8% more than comparable stand alone firms.

an alternative to centralized decision-making), or consider general coordination (or accommodation) decisions. Second, the idea that it may be efficient for the firm to have narrow scope and/or choose a boss that is biased toward particular workers is familiar from the work of Shleifer and Summers (1988), Rotemberg and Saloner (1994, 2000), and Van den Steen (2001). These papers emphasize the effect of narrow scope and bias on worker incentives rather than on private benefits or wages, but the underlying premise, that workers care about the boss's preferences, is the same. However, none of these papers analyzes firm boundaries. Third, there are several recent papers that like this one use the idea that some actions are noncontractible ex ante and ex post but may be transferable through owership; see, e.g., Aghion, Dewatripont and Rey (2002), Baker et al. (2002b), Bolton and Dewatripont (2001), Hart and Moore (2000), Holmstrom (1999), and Mailath et al. (2002). Mailath et al. (2002) is probably the paper closest to ours. They analyze a model in which the boss of an integrated firm maximizes total profit and so internalizes externalities. However, the cost of integration in their paper is a decline in worker initiative resulting from hold-up, rather than an insufficient emphasis on private benefits.

2. A basic model of coordination

Our first model concerns two units, which might operate in the same or similar output or input markets. Each unit makes decisions that affect the other unit. We will analyze whether the interaction between units will be regulated more efficiently if the units remain independent or if they integrate.[6]

Each unit has a binary decision: it can choose "Yes" or "No." We interpret Yes to mean that the unit acts in an "accommodating" (or compromising) way vis-a-vis the other unit, and No to mean that it does not. To have a concrete example in mind, think of two newspapers. Each one can decide whether to cater to the mass of readers or to specialize. Accommodation by one unit corresponds to the case where it specializes, for instance, by going up market. We focus on the binary case, because it highlights clearly the two main economic reasons for integrating.

---

[6]Here efficiency is defined with respect to the owners, managers and workers of the units. We ignore the impact of the units' decisions on consumers, e.g., in the form of reduced competition. Of course, in certain contexts, such "anti-trust" issues may be very important.

At the beginning an organizational form is selected--specifically, whether the units should be separate firms (nonintegration, i.e., there are two bosses) or should merge into one firm (integration, i.e., there is one boss), and also what kind of boss should be in charge. Next, each unit decides whether to choose Yes or No. Finally, the payoffs are realized. The time-line is as in Figure 1.

*	*	*
Organizational form chosen	Decision made	Payoffs realized

Figure 1

We assume that the decision in each unit is made by the boss of that unit. We also suppose that these decisions are noncontractible, both ex ante and ex post. This implies that contracts in which one unit agrees to choose Yes, say, in return for a side-payment from the other unit cannot be enforced. In other words, Coasian bargaining will not ensure efficiency of the accommodation decisions.

We represent the payoffs from different outcomes in the following matrix. We assume that these payoffs are nonverifiable and, for simplicity, perfectly certain.

Unit A		Unit B: Y	Unit B: N
	Y	A: $v_A(Y,Y)$, $w_A(Y,Y)$ B: $v_B(Y,Y)$, $w_B(Y,Y)$	A: $v_A(Y,N)$, $w_A(Y,N)$ B: $v_B(Y,N)$, $w_B(Y,N)$
	N	A: $v_A(N,Y)$, $w_A(N,Y)$ B: $v_B(N,Y)$, $w_B(N,Y)$	A: $v_A(N,N)$, $w_A(N,N)$ B: $v_B(N,N)$, $w_B(N,N)$

Figure 2

Here unit A is the row player and unit B is the column player. The entries are the payoffs. The first coordinate “v” refers to profit and the second coordinate “w” to private benefits (private benefits are measured in money). Subscripts refer to units, i.e., (v_A, w_A) denotes unit A’s payoffs and (v_B, w_B) unit B’s.

It will be convenient to introduce the notation

(2.1) $$z_A \,/\, v_A + w_A, \; z_B \,/\, v_B + w_B.$$

Here z_A refers to total surplus in unit A, z_B to total surplus in unit B, and $z_A + z_B$ equals aggregate social surplus.

As noted in the introduction, private benefits refer (broadly) to job satisfaction, or on-the-job consumption. It is reasonable to suppose that part of job satisfaction stems from the ability to pursue an independent course or agenda. Thus we will assume that coordination (accommodation by both units) leads to a reduction in total private benefits:

(2.2) $$w_A(Y,Y) + w_B(Y,Y) \# w_A(N,N) + w_B(N,N).$$

Note that we allow one side, but not both, to benefit privately from coordination. This may be reasonable, for instance, if coordination corresponds to the adoption of a common standard (see later). Also, note that we put no restrictions on whether coordination increases or decreases profits; moreover, even if coordination increases total profits, profits may rise by more or less than the fall in private benefits, i.e., (Y,Y) may be more or less (socially) efficient than (N,N) ($z_A(Y,Y) + z_B(Y,Y) \gtrless z_A(N,N) + z_B(N,N)$).

We mentioned in the introduction several interpretations of private benefits. We begin by focusing on the simplest one: Each unit contains one individual (a manager), who enjoys the private benefits w; in addition, a boss can divert all the profit from the units she operates. Then there are two leading organizational forms:

(1) Nonintegration: Manager A is the boss of unit A and manager B is the boss of unit B, i.e., the units are sole proprietorships. Manager A maximizes z_A and manager B maximizes z_B.

(2) Integration: A professional manager (outsider) is the boss of both units and managers A and B are subordinates. The professional manager maximizes $v_A + v_B$.

Note that, in (1), the managers play a noncooperative game, while in (2) no strategic elements are involved: the professional manager simply maximizes total profit.

Note also the critical role played by the assumption that the accommodation/non-accommodation decisions are noncontractible. In the absence of this, the parties would negotiate to an efficient ex post outcome using sidepayments, i.e., $z_A + z_B$ would be maximized under all organizational forms.

(1) and (2) are not the only possible organizational forms. If the units are under common ownership, manager A (resp., manager B), rather than a professional, could be the boss; she would then maximize $v_A + v_B + w_A$ (resp., $v_A + v_B + w_B$). Furthermore, even if the units are separately owned, they could be run by a professional manager (it is even possible that manager A could be the boss of unit B and vice versa). However, in the case of certainty, the first-best can always be achieved by (1) or (2). Thus we focus on these.

Although the decisions Yes or No are noncontractible ex ante and ex post, we will assume that organizational form is contractible at the beginning of the period. Coasian bargaining will then ensure that organizational form is chosen efficiently, i.e., to maximize total surplus ($z_A + z_B$) (there are no wealth constraints and information is symmetric). To spell this out a little more: if the units are initially manager-owned but integration is more efficient, then a professional manager will purchase control rights from managers A and B at prices such that all three parties are better off (in particular, manager i's private benefit under the professional manager plus the lump sum she receives for her control rights is at least equal to her payoff under nonintegration); and if the units are initially owned by a professional manager, and nonintegration is more efficient, then managers A and B will purchase control rights from the professional manager at prices such that all three parties are better off.

In summary, organizational form is chosen at the beginning of period 1 to maximize the value of $z_A + z_B$ in the subsequent game.

We will make the following technical assumption. We will suppose that total surplus and total profit are maximized either if both parties accommodate or if neither does. That is,

$$(2.3) \quad \text{Max}\,[z_A(Y,Y) + z_B(Y,Y),\, z_A(N,N) + z_B(N,N)] \\ \$\ \text{Max}\,[z_A(Y,N) + z_B(Y,N),\, z_A(N,Y) + z_B(N,Y)],$$

$$(2.4) \quad \text{Max}\,[v_A(Y,Y) + v_B(Y,Y),\, v_A(N,N) + v_B(N,N)] \\ \$\text{Max}\,[v_A(Y,N) + v_B(Y,N),\, v_A(N,Y) + v_B(N,Y)].$$

We will see that in this model nonintegration can cause two kinds of inefficiency. The first occurs when coordination yields greater social surplus than non-coordination, but one party loses out, i.e., the distribution of payoffs is uneven; this party will then refuse to accommodate. The second inefficiency occurs when both parties benefit from coordination but coordination is not a Nash equilibrium, i.e., at least one party wants to deviate (this is the standard prisoner's dilemma or free rider problem).

Pure coordination

It will be useful to separate out these effects even though in most real world situations they are combined. Thus we begin by analyzing the case where only the first inefficiency is present; we call this pure coordination:

$$(2.5) \qquad v_A(Y,N) = v_A(N,Y) = v_A(N,N), \qquad v_B(Y,N) = v_B(N,Y) = v_B(N,N),$$
$$w_A(Y,N) = w_A(N,Y) = w_A(N,N), \qquad w_B(Y,N) = w_B(N,Y) = w_B(N,N).$$

Pure coordination refers to a situation where accommodation by one party has no effect at all on payoffs, i.e., it "takes two to accommodate." An example would be if each party must decide whether to adopt a standard, and the standard has no force unless both parties adopt. Under pure coordination, the prisoner's dilemma or free rider problem does not arise since non-accommodation by either party is equivalent to non-accommodation by both.

Given (2.5), it is easy to see what the trade-off between nonintegration and integration is. Under nonintegration, each boss in effect has a veto on coordination, since by choosing non-accommodation she can achieve the (N,N) outcome. Thus (Y,Y) will occur if and only if (Y,Y) Pareto dominates (N,N):

$$(2.6) \qquad z_A(Y,Y) \geq z_A(N,N), \quad z_B(Y,Y) \geq z_B(N,N).$$

Note that even if (2.6) holds, (N,N) is always a Nash equilibrium along with (Y,Y); however, we will suppose that the parties do not pick a Pareto dominated equilibrium.

It is of course immediate that (2.6) implies that (Y,Y) is socially efficient:

$$(2.7) \qquad z_A(Y,Y) + z_B(Y,Y) \geq z_A(N,N) + z_B(N,N).$$

However, equally clearly, the converse is not true: (2.7) does not imply (2.6). This is the problem of winners and losers, mentioned earlier. Even though aggregate surplus may rise, the distribution may be such that one party loses out, and this party will then veto coordination, leading to the outcome (N,N).

Let us now turn to integration. Under integration, a single boss chooses the outcome that maximizes total profit. This outcome will be (Y,Y) or (N,N) according to whether

(2.8) $$v_A(Y,Y) + v_B(Y,Y) \gtrless v_A(N,N) + v_B(N,N).$$

It follows from (2.2) that, under integration, (Y,Y) is always the outcome if it is efficient, but may be the outcome even if it is inefficient.

Figure 3 illustrates the outcomes under nonintegration and integration.[7] In the figure, $\Delta v_i = v_i(Y,Y) - v_i(N,N)$ and $\Delta w_i = w_i(Y,Y) - w_i(N,N)$, i = A, B; these are the changes in profits and private benefits that result from a decision by both parties to accommodate. We have drawn the figure keeping Δw_A, Δw_B fixed and letting Δv_A, Δv_B be variable, but this of course is done merely to stay within two dimensions. In general, we are interested in the mapping from profits v and private benefits w into the choice of organizational form.

============

Figure 3 here

============

The first-best decision rule (2.7) is represented by the line FB-FB; all pairs of profit gains (Δv_A, Δv_B) that fall above this line call for coordination. In a similar way, line I-I represents the decision rule (2.8) and the quadrant NI-NI the decision rule under non-integration. Note that FB-FB is above I-I, because the sum of the changes in private benefits are assumed negative by (2.2).

The figure shows that nonintegration and integration make the opposite kind of mistake. Under nonintegration there is <u>too little</u> coordination in the sense that (Y,Y) never occurs if it is inefficient, but may not occur if it is efficient. Under integration, there is <u>too much</u> coordination in the sense that (Y,Y) is always the outcome if it is efficient, but may be the outcome even if it is inefficient. It follows that nonintegration achieves the first-best if (N,N) is efficient (since nonintegration errs on the side of too little coordination but never too much), while integration achieves the first-best if (Y,Y) is efficient (since integration errs on the side of too much

---

[7]We thank George Baker for suggesting this picture.

coordination but never too little).

It is also clear from Figure 3 that for some parameter values nonintegration and integration both achieve the first-best. This is the case when (Y,Y) is the outcome under nonintegration or when (N,N) is the outcome under integration. In all other cases it is essential to choose the right organizational form. Nonintegration will lead to inefficiency if profit gains from coordination are sufficiently unevenly distributed between the two firms; staying above the FB-FB line and moving in parallel with it will eventually take us outside the NI-NI acceptance region. Integration will lead to inefficiency if private costs from coordination are large enough; raising private costs will move the FB-FB/NI-NI part of the figure to the North-East and eventually make (N,N) the first-best choice.[8]

We summarize the discussion in

Proposition 1. Assume pure coordination, i.e., (2.5). Then, if (N,N) is efficient, nonintegration (with unit managers as bosses) achieves the first-best. If (Y,Y) is efficient, integration (with a professional manager as boss) achieves the first-best. Furthermore, integration will be uniquely optimal if profit gains from coordination are sufficiently unevenly divided between units. Nonintegration will be uniquely optimal if private costs from coordination are sufficiently high in aggregate.

General Coordination

The case of pure coordination is, of course, quite restrictive. In many situations, one party's decision to accommodate will affect payoffs, whether or not the other party accommodates. For example, accommodation could refer to a decision by one unit to raise its price (or lower its output), or not to poach on the unit's territory. If the other unit produces a

[8]In Figure 3 the region of inefficient integration could be reduced by letting manager B run the integrated unit. This way B's private benefits would be accounted for, something manager A would be happy with as long as she also dislikes coordination. However, (2.2) admits the case where A enjoys private benefits from coordination (say, because it involves the use of her firm's technology). In that case, (2.2) implies that manager B would coordinate too little relative to first-best and hence the comparison to a professional manager would be ambiguous. We do not explore these options here, since we get first best either way.

substitute good, this will increase the other unit's profit, regardless of whether it raises its price, or decides not to poach too.

We show in the Appendix that our analysis generalizes to cases where it does not "take two to accomodate," as long as we are prepared to make some assumptions.

An alternative interpretation of private benefits

So far we have emphasized the interpretation where the private benefits w_i are enjoyed by a single manager in each unit, who is the boss under nonintegration and a subordinate under integration. However, as noted in the introduction, there is a second interpretation of the model that is also of interest. We can imagine that the private benefits refer to the job satisfaction of unit i workers and that some bosses have goals that are (partially) congruent with those of the workers, i.e., they care about the same things. In particular, suppose that there are three types of bosses, unit A enthusiasts with preferences $m + \lambda_A w_A$; unit B enthusiasts with preferences $m + \lambda_B w_B$; and professional managers with preferences m. Here m is money and λ_A, λ_B represent congruence between a boss's goals and those of unit A and B workers.

Suppose now that a boss can divert a fraction $\theta < 1$ of total profit toward herself; one can imagine that she uses this profit for perks or empire-building activities that benefit her alone--fancy offices, secretaries, pet projects, etc. Denote profit by v. Then a unit i enthusiast will maximize

(2.9) $$\theta v + \lambda_i w_i,$$

while a professional manager will maximize v.[9]

If we make the simplifying assumption that $\theta = \lambda_A = \lambda_B$, this yields the same objective function for the different kinds of bosses as in the model without workers described above.[10]

[9]With partial diversion, it may be desirable to give manager A a share of the residual profits of unit B in order to get closer to first-best, but we will not pursue this possibility here.

[10]It is obviously important that there is no boss who is an enthusiast for <u>both</u> units A and B. Under integration such a boss would maximize $\theta(v_A + v_B) + \lambda_A w_A + \lambda_B w_B$, and, if $\theta = \lambda_A = \lambda_B$, this would yield the first-best.

In this interpretation, the choice of organizational form is made by the initial owner(s) of units A and B at the beginning of the period. They must decide whether the units should be separate (nonintegration) or together (integration), and what kind of boss to put in charge. Assume that owners face a competitive labor market and wages are agreed up front. Then the (total) wage ω_i for unit i workers will satisfy

(2.10) $$\omega_i + w_i = U,$$

where U is the (total) market clearing reservation wage and w_i refers to (expected) worker private benefits. Suppose that the initial owner(s) wish to sell out and retire, i.e., they are interested only in money. Then, given that side-payments are possible, organizational form will be selected to maximize the total value of the two units net of wages, i.e., given (2.10), $v_A + v_B + w_A + w_B$ (we ignore the private benefits and remuneration of bosses).

We may conclude that the analysis of this second interpretation of the model is identical to the previous one.

Discussion

We have interpreted the model in this section as being about "horizontal interaction", suggesting that it is primarily about interactions between firms in the same market. This is needlessly narrow. Our approach has potential relevance whenever firms have to make decisions unilaterally, because bargaining over those decisions is too costly. Here we want to discuss briefly the broader relevance of the model using two illustrative examples.

Knowledge transfers. Knowledge is notoriously difficult to sell. It is often difficult to codify. It is also hard to circumscribe the use of knowledge once it has been transferred to another party. The transfer of knowledge from one firm to another may be very valuable, but it may be very difficult to arrange things so that the benefits of the transfer are appropriately divided between the firms. For this reason, knowledge transfer can be a major impetus for integration. The 1960s saw a wave of mergers and acquisitions, which partly were driven by regulation and partly were guided by new management technology that was best appropriated by buying up firms and implementing the new management systems after that.[11]

Note that once the knowledge transfer is complete and the new systems are up and running, one may well expect divestment of some of the units. Subsequent divestments are often read as signs of failed mergers, but if the initial reason for a merger was knowledge transfer, that verdict may be misguided. To study this sequence of events, we could extend our model to two periods. In the first period the key decision is whether to transfer knowledge by merging. In the second period, assuming a merger, divestment may become optimal, because decisions with significant private benefits come to dominate the payoffs. The analysis of such a two-period model is more complicated than piecing together two one-period analyses, because the payoffs from the second stage will determine how much should be paid in the first-period merger. Also, the first-period decision will in general influence the second period payoffs and options.

Technology choice. Standards are important in emerging industries like information technology. The social benefits from a common standard can be huge, because it increases the overall market. Yet, getting parties to agree to a standard is very difficult. Most of the time industry agreements have to rely on self-enforcement, which tends to be ineffective. Instead, proprietary systems may become de facto standards. Cisco's operating system is an example of a dominant standard/platform in the Internet space. Should an entrepreneur with a new idea choose to develop it for the Cisco standard or should it try to do something more generic? The entrepreneur needs to be assured that Cisco will not take advantage of his choice later on. If this is difficult contractually, as it often is, the alternative may be integration.

This situation again fits our model. Suppose the entrepreneur's total payoff goes down if he unilaterally chooses to choose Cisco's standard, because contractual guarantees are ineffective. Then, if the social value of using Cisco's standard is sufficiently positive, Cisco could buy out the entrepreneur. The transfer of decision rights would justify the transfer of payments. In the late 1990s Cisco was very active buying start-ups as a way of expanding its technology base. Cisco's reputation for preserving the entrepreneur's "private benefits" inside the firm made this strategy work particularly well (until the downturn hit).

---

[11]Much of the knowledge transfer is conducted by management consulting firms these days. The outsourcing model in the next section could be used to study that development.

A richer analysis would include an innovation stage as well as a buy-out stage. It seems reasonable to assume that the entrepreneur's private benefits are initially high relative to the benefits of a buy-out.[12] Indeed, one typically does not know whether the entrepreneur will be successful and if so, whether the innovation will fit Cisco's technology rather than some other company. (An important reason why private benefits are so high initially is that the entrepreneur has exaggerated beliefs about the probability of success; see, e.g. Van den Steen, 2001). Thus, it is often best to let innovation take place on a stand-alone basis. Once the project has matured to a stage where valuations are more clearly in view a buy-out may be optimal.

It is worth noting that dynamic extensions of the basic model would often lead to analyses that resemble those encountered in traditional models of hold-up. In the Cisco case the entrepreneur may not want to specialize to Cisco's standard, because he fears that Cisco will then by virtue of controlling the technology be able to appropriate much of the later benefits of cooperation. There are important differences, however, between our approach and the traditional way of modeling hold-ups using incomplete contracts. In the traditional model, individuals make noncontractible human capital investments; asset ownership influences the payoffs from these investments, but not who makes the investments. Our focus is not on human capital investments, typically, but on strategic decisions with opportunity costs for the organization; asset ownership determines who makes these decisions. This structure reflects the reality that firms are organizational entities rather than mere entrepreneurs. It is also a potentially simpler way of studying hold-up problems.

3. An extension to outsourcing.

In Section 2 we analyzed a model in which two units may find it easier to coordinate behavior (achieve accommodation) under common ownership than under separate ownership. In that model there was no role for outsourcing: decisions were always made inside the units and the only question was whether they were made by one boss or two.

---

[12] It may, of course, also be more important at this stage to motivate the entrepreneur to work hard.

In reality, outsourcing of a decision or activity is often a viable alternative for an organization. For example, consider two units, operating in the same or similar output markets, that need a specialized input. The input could be an intermediate good or it could be a service like advertising or marketing. Each unit can remain separate and provide the service in-house; or the two units can merge horizontally and collectively provide the service in-house; or the units can remain separate and procure the service from an independent supplier. In this section we will develop a model that allows us to analyze the trade-off between these choices (and others too). Among other things the model will throw light on the often-mentioned idea that firms merge horizontally to exploit "synergies"; here the synergies will stem from the ability to reduce costs by sharing scarce input capacity.

In order to allow for the possibility of outsourcing, we need to permit <u>some</u> contractibility; ie., we have to relax the assumption of Section 2 that everything is ex post noncontractible. Instead we will suppose that the parties can contract on the quantity and quality of input at the time of delivery. However, we will assume that at an earlier stage a decision must be made as to how specialized the input is. This decision is noncontractible (ex ante and ex post) and is made by the boss of the supplying unit. To make the trade-offs interesting, we also need the units' profits to be stochastic ex ante; that is, we will depart from the assumption of perfect certainty made in Section 2.

The economy lasts for one period. At the beginning of the period, organizational form is chosen. In the middle of the period any uncertainty is resolved and a specialization decision is made. Finally, at the end of the period trade occurs and payoffs are realized. The time-line is as in Figure 4.

*____________________*____________________*

Organizational form chosen	State learned and specialization decision made	Trade and payoffs realized

Figure 4

One buyer.

We start with the simplest case of one buyer operating in the output market and one seller operating in the input market--we refer to them as B and S. B requires a specialized input, one widget, say, at the end of the period. S has one unit of capacity, costing k, which can be used to supply one widget at the end of the period. S's variable costs are zero. This widget can be supplied either to B or to the outside market, which is assumed to be competitive. The competitive market price is r.

S must make a choice in the middle of the period. To supply B at the end of the period, S must specialize to B; but then S can't supply the outside market. Alternatively, S can choose to "remain flexible" and supply the outside market; but then S cannot supply B.

The value to B of a (specialized) widget is v. In addition B's manager receives a private benefit w if the widget is supplied. (We ignore any private benefits of S's manager or managers in the outside market; we assume they are independent of transactions with B.) The variables v, w, r are uncertain at the beginning of the period. However, the uncertainty is resolved before the specialization decision is made; moreover, v, w, r are observable to B and S (but are not verifiable). We assume that no long-term contracts can be written about the end-of-period widget price (because widget characteristics cannot be specified in advance) although spot contracts at the end of the period are possible.

As in Section 2, we suppose that the boss of a unit diverts all the profit from that unit. We also suppose that all parties are risk neutral.

In this model, the key decision is whether to specialize. This decision is made by whoever is the boss of S. We focus on two organizational forms:

(1) B and S are independent firms and each is run by its manager (nonintegration).

(2) B and S are a single firm run by B's manager (vertical integration).

As in Section 2, these are not the only possibilities. For example, B and S could be separate and B could be run by a professional manager. Or B and S could be one firm run by S's manager. But (1) and (2) are leading cases.

Assume first that B and S are nonintegrated. Suppose S specializes to B. Then at the end of the period S and B will bargain about the price of the input. Assume that they divide the gains from trade 50 : 50. Since B's boss values the widget at $v + w$ and S's (variable) costs are zero, this means that S will receive $\frac{1}{2}(v + w)$. In contrast, if S does not specialize and sells on the open market, S will receive r. It follows that S will specialize to B if and only if

(3.1) $$\tfrac{1}{2}(v + w) > r.$$

(Recall that S knows v, w, r when she decides whether to specialize.) In other words, we have a classic hold-up problem.

Now suppose B and S are integrated. Then the specialization decision is made by the boss of B, who maximizes total profit plus her private benefits w. So specialization will occur at date 1 if and only

(3.2) $$v + w > r.$$

Of course, this is also the first-best rule. We conclude that in the simple case where there is only one buyer and one seller integration is superior to nonintegration. The total net surplus under integration is given by

$$E \operatorname{Max}(v + w, r) - k,$$

where the expectation is taken with respect to the probability distribution of (v, w, r).

Two buyers

Let us consider next what happens if we have two buyers in the output market: call them B1, B2. We assume that a second unit of capacity is available in the input market at the same cost k as the first unit. A key question is whether there should be two units of capacity or just one (in which case supply is obviously available ex post to only one buyer). We already know that conditional on two units of capacity being available the first-best can be achieved by having B1, B2 each vertically integrate with a supplier. This is illustrated in Figure 5 (i). On the other hand, if there is only one unit of capacity, then there are two leading organizational forms, also illustrated in Figure 5. In (ii), B1, B2 and a single S all merge. In (iii), B1, B2 and a single S all stay separate

============

Figure 5 here

============

.

As in Section 2, an important aspect of organizational form concerns the identity of the boss. To simplify matters we will suppose that in (ii) B1, B2 and S are run by a professional manager, while in (iii) each of B1, B2 and S is run by its own manager. Note also that there are two other "non-leading" organizational forms. First, B1 and B2 may merge horizontally with S staying independent. Second, B1 and S may merge vertically, with B2 staying independent (or vice versa). We will return to these two forms below.

We suppose that Coasian bargaining ensures that organizational form is chosen at the beginning of the period to maximize the expected value of total surplus. Implicit in this assumption is the idea that organizational form, and capacity, are contractible. That is, if, say (i) is optimal, then B1 or B2 cannot (separately) sell off some of their capacity after organizational form has been determined. And if, say, (iii) is optimal, one of the buyers cannot merge (secretly) with S.[13]

To understand the trade-offs between (i) - (iii), it is useful to work with the following example. Suppose both buyers have the same private valuations, denoted $w = w_1 = w_2$. Assume w and the outside value r are constants, while v can take on two values: $v = v_H$ with probability π and v_L with probability $(1 - \pi)$. Assume also that $v_L + w > r$, so that it is always efficient to supply B1 and/or B2 rather than the outside market. Then, with two units of capacity, first-best (expected) gross surplus is

[13]An alternative assumption is made in (part of) Bolton and Whinston (1993). We return to this in footnote 14.

(3.3) $$Z^{**} = 2\{\pi(v_H + w) + (1 - \pi)(v_L + w)\}$$

since each buyer will always receive a widget (whose value may be v_L or v_H). On the other hand, with one unit of capacity, first-best (expected) gross surplus is

(3.4) $$Z^* = (2\pi - \pi^2)(v_H + w) + (1 - \pi)^2 (v_L + w),$$

since the single widget will be supplied to the buyer with value v_H if there is one, and the probability that at least one B has $v = v_H$ is $1 - (1 - \pi)^2 = 2\pi - \pi^2$.

The first thing to notice is that

(3.5) $$\begin{aligned} Z^{**} - Z^* &= 2\{\pi(v_H + w) + (1 - \pi)(v_L + w)\} - \{(2\pi - \pi^2)(v_H + w) + (1 - \pi)^2 (v_L + w)\} \\ &< (2\pi - \pi^2)(v_H + w) + (1 - \pi)^2 (v_L + w) = Z^*, \end{aligned}$$

i.e., there are diminishing returns to capacity. The right-hand side represents the marginal gain from the first unit of capital and the left-hand side the marginal gain from the second unit. The reason for the diminishing returns is that with one unit of capital, winner-picking is possible (cf. Stein (1997)): the scarce input can be directed to where it is most needed; moreover, with positive probability B1 and B2 do not have strong needs (v_H) at the same time. It follows that, if $k < Z^{**} - Z^*$, two units of capacity are first-best optimal, while, if $Z^{**} - Z^* < k < Z^*$, one unit of capacity is first-best optimal. Finally, if $k > Z^*$, it is first-best optimal to close down (i.e., to have zero units of capacity).

Consider now the second-best. We know that, if two units of capacity are first-best optimal, then since symmetric vertical integration achieves the first-best, two units of capacity are also second-best optimal. Similarly, if close-down is first-best optimal, then close-down is second-best optimal.

The interesting case is where $Z^{**} - Z^* < k < Z^*$, i.e., it is first-best optimal to have one unit of capacity. Under these conditions the first-best is no longer generally achievable since forms (ii) and (iii) may each be inefficient. We analyze them in turn.

Form (ii)

Form (ii) may be inefficient because the single merged firm is run by a professional manager. The manager, since she maximizes total profit but ignores private benefits, adopts the following specialization rule:

$$\begin{aligned}
&\text{Specialize to B1 if } v_1 = \text{Max}(v_1, v_2) \geq r,\\
(3.6)\qquad &\text{Specialize to B2 if } v_2 = \text{Max}(v_1, v_2) \geq r,\\
&\text{Don't specialize if Max}(v_1, v_2) < r,
\end{aligned}$$

where v_1, v_2 are respectively B1, B2's valuations. In contrast, the first-best specialization rule is:

$$\begin{aligned}
&\text{Specialize to B1 if } v_1 + w = \text{Max}(v_1 + w, v_2 + w) \geq r,\\
(3.7)\qquad &\text{Specialize to B2 if } v_2 + w = \text{Max}(v_1 + w, v_2 + w) \geq r,\\
&\text{Don't specialize if Max}(v_1 + w, v_2 + w) < r.
\end{aligned}$$

We see that under (ii) the professional manager specializes too little because she ignores the private benefit w. However, conditional on specializing, she gets the balance between B1 and B2 right: she specializes to B1 if and only if $v_1 \geq v_2$, that is, whenever $v_1 + w \geq v_2 + w$.

Form (iii)

Form (iii) has the advantage over form (ii) that B1 and B2's managers internalize the private benefits w. However, form (iii) has the disadvantage that there is a hold-up problem. S's manager recognizes that if she specializes to B1 or B2 she will be locked into this particular buyer and will split the surplus with them. Hence, S adopts the following rule:

$$\text{Specialize to B1 if } v_1 \geq v_2 \text{ and } \tfrac{1}{2}(v_1 + w) \geq r,$$

(3.8) Specialize to B2 if $v_2 \$ v_1$ and ½ $(v_2 + w)$ $\$r$,

Don't specialize if ½ Max $(v_1 + w, v_2 + w) < r$.

It follows that under (iii) there will again be too little specialization relative to the first-best, but this time because $(v_i + w)$ is multiplied by the coefficient ½.. Note also that S again gets the balance between B1 and B2 right: she specializes to B1 if and only if ½ $(v_1 + w) \$$½ $(v_2 + w)$, that is, whenever $v_1 + w \$ v_2 + w$.

We see that forms (ii) and (iii) may both be inefficient, but in different ways. We now present examples showing that, depending on the circumstances, (ii) can be better than (iii) or vice versa. In fact, in the first example, (ii) achieves the first-best, while in the second example (iii) achieves the first-best. Hence these examples also show that (ii) and (iii) can both be second-best optimal organizational forms.

<u>Example 1</u>: $v_H = 12$, $v_L = 8$, $\pi = ½$, $w = 0$, $r = 7$.

This example is very simple given that $w = 0$. It is clear from a comparison of (3.6) and (3.7) that (ii) achieves the first-best conditional on one unit of capacity being optimal. However, we know that one unit of capital is indeed first-best optimal if $Z^{**} - Z^* < k < Z^*$, i.e., if $9 < k < 11$.

It only remains to check that (iii) is suboptimal when $9 < k < 11$. But this follows from that fact that, since ½ $(v_H + w) < r$, S won't be prepared to specialize to either B1 or B2 when S is independent.

<u>Example 2</u>: $v_H = 9$, $v_L = 4$, $\pi = ½$, $w = 7$, $r = 5$.

Now, conditional on one unit of capacity being used, outsourcing (iii) is superior to complete integration (ii). The reason is that since ½ $(v_L + w) > r$, an independent supplier will be prepared to specialize to either buyer, and will obviously prefer to specialize to the buyer with the higher v, i.e., (iii) achieves the first-best. In contrast, complete integration (ii) is inefficient since the boss will direct input to the outside market when both buyers have low v's, since $v_L < r$.

The final point to note is that one unit of capacity is indeed first-best optimal if Z** - Z* = 13½ < k < 19¾ = Z*.

Examples 1 and 2 are special in that either (ii) or (iii) achieves the first-best. Typically, this will not be so. The general situation is illustrated in Figure 6.

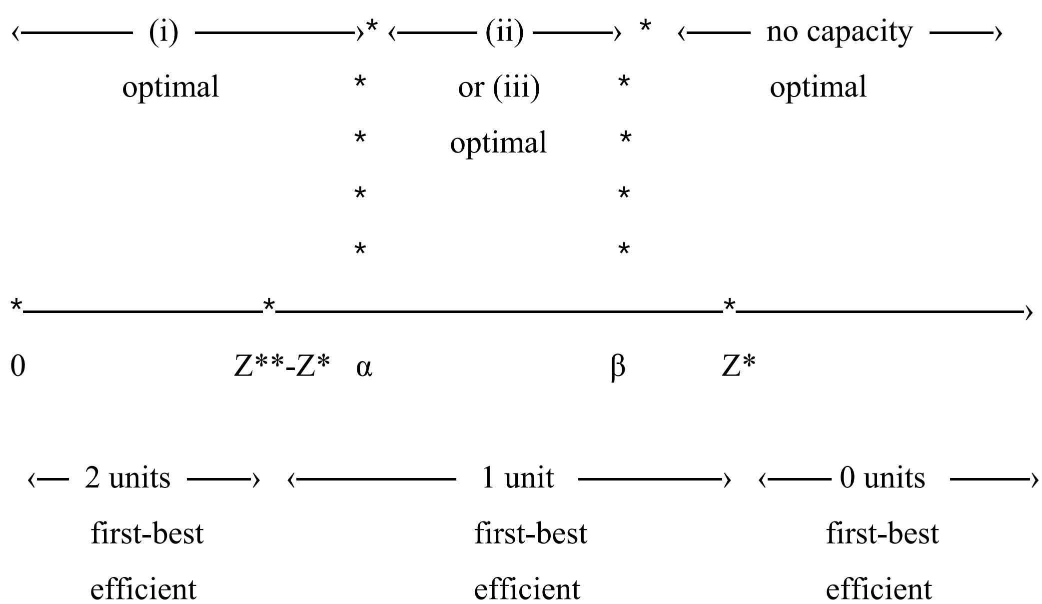


Figure 6

To understand the figure, note that in the second-best the comparison between forms (ii) and (iii) depends on the probability distribution of (v, w, r), but not on k, because both (ii) and (iii) are conditional on one unit of capacity being employed. Let Z be the gross surplus corresponding to the better of these two forms. Then the optimal capacity in the second-best depends on a comparison between Z - k, Z** - 2k and 0. Define

$$\alpha = \text{Min}\ \{Z^{**} - Z,\ Z^{**}/2\}, \quad \beta = \text{Max}\ \{Z, Z^{**}/2\}.$$

Then 2 units of capacity are second-best optimal if k < α, 1 unit is optimal if α < k < β and 0 units are optimal if k > β. (The role of Z**/2 in the definitions of α and β is to make sure that α is less than β.)

We know that $Z \# Z^*$, because it is second-best. Therefore, using (3.5), $\alpha \; \$ \; Z^{**} - Z^*$ and $\beta \# Z^*$, implying that the region in which one unit of capacity is optimal is no larger in the second-best than in the first-best and strictly smaller whenever $Z < Z^*$, as drawn in Figure 6. (If $\alpha = \beta$ the middle region will disappear altogether.)

Other organizational forms

Let us now consider the two "non-leading" organizational forms mentioned earlier. In the first, B1 and B2 merge horizontally, with S staying independent. In the second B1 and S merge vertically, with S staying independent (or vice versa).

The first of these can easily be ruled out (at least if the boss of B1/B2 is a professional manager). Under a B1/B2 merger, the specialization rule becomes:

Specialize to B1 if $\frac{1}{2} v_1 = \frac{1}{2} \text{Max}(v_1, v_2) \; \$ \; r$,
(3.9) Specialize to B2 if $\frac{1}{2} v_2 = \frac{1}{2} \text{Max}(v_1, v_2) \; \$ \; r$,
Don't specialize if $\frac{1}{2} \text{Max}(v_1, v_2) < r$.

It is always better to go the extra step and include S in the merger. Under complete integration (form (ii)) the specialization rule is:

Specialize to B1 if $v_1 = \text{Max}(v_1, v_2) \; \$ \; r$,
(3.10) Specialize to B2 if $v_2 = \text{Max}(v_1, v_2) \; \$ \; r$,
Don't specialize if $\text{Max}(v_1, v_2) < r$.

(3.9) and (3.10) both lead to two little specialization relative to the first-best (3.7), but (3.10) gets closer to the optimum. Thus a complete B1/B2/S merger dominates a partial B1/B2 merger.

The second "non-leading" organizational form is where B1 and S merge vertically, with B2 staying independent ("asymmetric" vertical integration). This leads to the following specialization rule:

$$\text{(3.11)}\quad \begin{array}{l} \text{Specialize to B1 if } v_1 + w \geq \tfrac{1}{2}(v_2 + w),\ v_1 + w \geq r, \\ \text{Specialize to B2 if } v_1 + w < \tfrac{1}{2}(v_2 + w),\ \tfrac{1}{2}(v_2 + w) \geq r, \\ \text{Don't specialize otherwise.} \end{array}$$

The reason is that the boss of B1/S puts value $v_1 + w$ on a widget sold to B1, but only value $\frac{1}{2}(v_2 + w)$ on a widget sold to B2, given that half the surplus goes to B2.

Relative to the first-best (3.7), we see that the boss of B1/S makes the socially correct decision concerning whether to supply B1 or the outside market (she compares $v_1 + w$ to r), but that she will be biased in favor of specializing to B1 rather than B2 (she compares $v_1 + w$ to $\frac{1}{2}(v_2 + w)$).

In some cases, the first effect may be important enough that asymmetric vertical integration is indeed optimal. However, if

$$\text{(3.12)}\qquad \tfrac{1}{2}(v_H + w) < v_L + w,$$

then the boss of B1/S will <u>never</u> specialize the widget to B2 since specialization is unprofitable even when B2's value is high and B1's is low. As a result asymmetric vertical integration is equivalent to vertical integration between B1 and S with B2 absent (which yields net social surplus $\pi_H (v_H + w) + \pi_L (v_L + w) - k$); but this in turn is dominated either by zero capacity (if $k > \pi_H (v_H + w) + \pi_L (v_L + w)$) or by symmetric vertical integration, i.e., form (i) (if $k < \pi_H (v_H + w) + \pi_L (v_L + w)$).

Discussion

We conclude this section by mentioning a piece of empirical evidence that is consistent with our analysis, and by relating our model to that of Bolton and Whinston (1993). First the evidence. There are several studies that show that, when a sector is in decline, firms divest weak units and these are often sold to other firms in the same line of business (see the references in Meyer et al. (1992)). Meyer at al. offer an influence cost explanation for this phenomenon. Our model provides an alternative explanation. A fall in a sector's prospects can be represented by an increase in the cost of capacity k. Consider Figure 6. Suppose that we make the plausible assumption that ½ $(v_H + w) < r$, i.e., the hold-up problem is sufficiently severe that specialization never occurs under outsourcing. Then form (ii) is superior to form (iii). It follows that as k rises the economy will move from a region in which symmetric vertical integration is optimal to one where complete integration is optimal. That is, it becomes advantageous for the two buyers to save costs by sharing scarce input capacity, and the way for them to do this is to merge horizontally (they are already integrated vertically with their suppliers). In effect, there is industry consolidation.

Consider next the comparison with the work of Bolton and Whinston (1993) (henceforth BW). BW is one of the first papers to analyze integration decisions in a situation where several buyers are supplied input by a seller S that may have scarce capacity. BW consider a situation where there are two buyers B1 and B2 in a downstream market and a single seller S that with probability $(1 - \lambda)$ has the capacity to supply both, and with probability λ can supply only one. BW employ a traditional property rights model in which bargaining leads to ex post efficiency and the only distortions are in the (nontransferable) ex ante investments by managers of B1, B2 (the manager of S does not invest). With this setup BW analyze all the organizational forms that we do, with the exception of symmetric vertical integration (for most of the paper, they consider only one seller).[14]

BW show that nonintegration (form (iii)), complete integration (form (ii)), and asymmetric vertical integration (one of our "nonleading" forms) can all be optimal. NI is efficient if λ • 1, since, given their assumption of outside option bargaining, the downstream firm that receives the scarce input pays a price equal to the value of the input to the other downstream firm; downstream firms are therefore not held up by S and invest efficiently. Complete integration may be efficient if $\lambda < 1$, since although one downstream manager will underinvest given that he will be held up by the owner of the single firm, the other downstream manager, if he is the owner, will invest efficiently. Asymmetric vertical integration has the advantage over complete integration in

[14]As we pointed out in footnote 13, BW also consider the case where organizational form is not contractible. They show that, when nonintegration is socially optimal, it may pay one of the buyers secretly to integrate vertically with S. Although we do not consider this possibility here, our analysis could be modified to do so.

that the manager of the nonintegrated downstream firm will underinvest less; but the disadvantage in that the manager of the integrated downstream firm will overinvest, given that this increases his bargaining power with respect to the nonintegrated downstream firm. Finally, a horizontal merger by B1 and B2 is not optimal since this allows S to hold up both downstream firms.

While several of our results are similar to those of BW (like us they find that forms (ii) or (iii) may be optimal, but are able to rule out a horizontal merger by B1/B2), there are also significant differences. First, and most obviously, we focus on how organizational form affects the trade-off between private benefits and profits, rather than on the trade-off between different managerial investments. Second, we endogenize capacity choice, so that symmetric vertical integration is a possible organizational form. Third, with a not unreasonable assumption, we are able to rule out asymmetric vertical integration. Fourth, we believe that our analysis is simpler.

Perhaps the most important difference is that we are able to capture the idea that firms may merge horizontally to exploit "synergies." BW's framework is not able to incorporate the idea of a synergy since, given their assumption that all variables except for nontransferable investments are ex post contractible, bargaining leads to ex post efficiency under any organization form, i.e., the only inefficiencies are in ex ante managerial investments.

4. Conclusion.

In the traditional property rights model investments are inalienably attached to individuals; asset transfers change individual incentives to invest, but not who makes the investments. In our model decisions, including investment decisions, are transferable through ownership. Our structure is in many ways close to the traditional view of the firm as a technologically defined entity that makes decisions about inputs, outputs and prices. The difference is that our firm does not necessarily maximize profits, because it may be desirable to choose a boss that cares about nontransferrable private benefits. It is this small wrinkle in the traditional model that opens the door to a discussion of boundaries. In the traditional model, barring regulatory constraints, it would be optimal to organize all activity in a single firm.

The fact that our view of the firm is so close to the traditional view should make our approach easier to apply to questions of industrial organization. The model in Section 3 on outsourcing is indicative of this. One could pursue a similar analysis with a larger number of firms to understand industry structure. One could also study industry dynamics – times of entrepreneurial growth followed by times of industry consolidation, for instance.

Our framework may also be helpful for analyzing the internal structure of the firm. In a previous version of the paper, we looked at a simple model of delegation and we plan to return to it in future research. We are interested in understanding how the internal organization of the firm influences and is influenced by its boundaries. Delegation creates new, hybrid opportunities for matching decision making objectives with decision making authority. How do the firm's de facto objectives change with delegation and how do additional activities influence what gets decided – these are examples of questions we would like to understand better. Because our approach emphasizes the role of the firm's objective function it would also be of interest to analyze alternative governance structures for decision making, such as cooperatives.

Private benefits play a pivotal role in our analysis. It could have been more natural to let incentive effects counter-balance the benefits from integration (as in Mailath, et al, 2002). We chose the current path, because the framework seems more flexible and the analysis more tractable. It remains to be seen whether private benefits can be defined tightly enough and are empirically important enough for the approach to be useful.

One of our objectives in writing this paper has been to move the focus of attention away from assets towards activities. Asset ownership is at the core of the property rights theory and it will remain important for understanding boundaries. At the same time it is remarkable how few practitioners, organizational consultants and researchers studying organizations within other disciplines than economics (e.g. sociology and organizational behavior) ever talk about firms in terms of asset ownership. For most of them a firm is defined by the things it does and the knowledge and capabilities it possesses. Coase (1988) in his article "Industrial Organization: A Proposal for Research," makes clear that he too is looking for "a theory which concerns itself with the optimum distribution of activities, or functions, among firms." (p 64). He further notes "the costs of organizing an activity within any given firm depend on what other activities the firm is engaged in. A given set of activities will facilitate the carrying out of some activities but hinder the performance of others."

The model we have proposed is in this spirit.

APPENDIX: *General coordination.*

Here we show that the main conclusions of Proposition 1, concerning pure coordination, hold more generally. From now on we drop (2.5), but replace it by:

(A.1) $z_A(N,Y) \$ z_A(N,N), z_B(Y,N) \$ z_B(N,N),$

(A.2) $z_A(N,N) \$ z_A(Y,N), z_B(N,N) \$ z_B(N,Y).$

Condition (A.1) says that, under nonintegration, each boss is better off if the other boss is accommodating, given that she does not accommodate. Condition (A.2) says that, under nonintegration, it never pays a boss to be accommodating if the other boss is not accommodating.

Note that these assumptions are implied by (2.5), i.e., they hold under pure coordination. Also, they are plausible in the case where the two units produce substitute goods and accommodation refers to a decision by one unit to raise price or not poach on the other unit's territory.

Given (A.1) - (A.2), we can establish the following proposition.

Proposition 2: Assume (A.1) - (A.2). Then:

(A) (N,N) is always a Nash equilibrium in the game in Figure 2.

(B) Suppose (N,N) is efficient. Then all Nash equilibria generate the same payoff vector $(z_A(N,N), z_B(N,N))$.

(C) Suppose (Y,Y) is efficient. Then, if

(A.3) $z_A(Y,Y) \$ z_A(N,Y)$ and $z_B(Y,Y) \$ z_B(Y,N)$,

(Y,Y) is a Nash equilibrium and Pareto dominates all other Nash equilibria. On the other hand, if (A.3) is violated, all Nash equilibria are payoff equivalent to (N,N).

As before, we will assume that, if there are multiple equilibria, the parties will never pick a Pareto dominated equilibrium. Thus we can rephrase Proposition 2 as follows. Under nonintegration, if (N,N) is efficient, the outcome will be (N,N). In contrast, if (Y,Y) is efficient, the outcome will be (Y,Y) if (A.3) holds but (N,N) otherwise.

Proof of Proposition 2: Part (A) is obvious; (N,N) is always a Nash Equilibrium given (A.2).

To establish (B), note that, if (Y,Y) yields strictly less surplus than (N,N), either $z_A(Y,Y) < z_A(N,N)$ or $z_B(Y,Y) < z_B(N,N)$. But then, given (A.1), either boss A or boss B – say, boss A for concreteness – will deviate from (Y,Y) and so (Y,Y) cannot be a Nash equilibrium. (The argument is a little more complex if (Y,Y) yields the same surplus as (N,N).) Furthermore, given (A.1) - (A.2), N is a dominant strategy for boss A. Suppose (N,Y) is a Nash Equilibrium. Then (A.2) implies $z_B(N,N) = z_B(N,Y)$ and, because (N,N) is efficient, $z_A(N,N) = z_A(N,Y)$ by (A.1). So, (N,Y) yields the same payoffs as (N,N). A parallel argument establishes that if (Y,N) is a Nash Equilibrium it must be payoff equivalent to (N,N). This proves part (B).

In Case (C), the game is supermodular, and so by standard results for such games (see Milgrom and Roberts, 1990), (Y,Y), being efficient, Pareto dominates all other Nash equilibria. If (A.3) is violated, (Y,Y) is not a Nash Equilibrium and the argument proceeds as in Case B.

Q.E.D.

Like Proposition 1, Proposition 2 says that nonintegration achieves efficiency if (N,N) is efficient, but may not achieve it if (Y,Y) is efficient. However, there are now two reasons why nonintegration may not achieve (Y,Y) when it is efficient (as opposed to the single reason in the case of pure coordination). First, as in the case of pure coordination, the benefits from coordination may be so unevenly distributed that (Y,Y) does not Pareto dominate (N,N). That is, we may have $z_A(N,N) > z_A(Y,Y)$ or $z_B(N,N) > z_B(Y,Y)$, which, given (A.1), implies that (A.3) is violated. Second, even if (Y,Y) Pareto dominates (N,N), (A.3) may not hold anyway, i.e., a unilateral deviation from coordination may be in the interest of one party, even though it is self-defeating (it leads to the Nash equilibrium (N,N) that both parties dislike).

Given this second reason, the characterization of the outcome under nonintegration is not quite as simple as in the case of pure coordination. Specifically, the set of parameter values under which nonintegration yields (Y,Y) is now a subset of the NI-NI quadrant in Figure 3, rather than the whole quadrant. Note, however, that this merely reinforces the result that nonintegration errs on the side of too little coordination.

The analysis of integration is unchanged under (A.1) - (A.2), since the assumption of pure coordination (2.5) was never used in this case. Since, as in the case of pure coordination, nonintegration errs on the side of too little coordination, while integration errs on the side of too much, we have established

Proposition 3. Assume (A.1) - (A.2). Then, if (N,N) is efficient, nonintegration achieves the first-best. On the other hand, if (Y,Y) is efficient, integration achieves the first-best.

We may conclude that, given the (admittedly very strong) assumption of perfect certainty and binary decisions, the main results of the pure coordination case continue to hold for more general kinds of decisions.

REFERENCES.

Aghion, P., M. Dewatripont and P. Rey (2002), "Transferable Control," mimeo, University College London.

Baker, G., R. Gibbons and K. Murphy (2002a), "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics*, 117(1):39-84.

________________ (2002b), "Relational Contracts in Strategic Alliances", working paper, MIT.

Baker, G. and T. Hubbard (2002) "Contractibility and Asset Ownership: On-Board Computers and Governance in U. S. Trucking," *American Economic Review*, forthcoming.

Blair, M.. and L. Stout (1999), "A Team Production Theory of Corporate Law," *Virginia Law Review*, 85(2).

Bolton, P. and M. Dewatripont (2001), *Introduction to Contract Theory*, forthcoming, MIT Press.

Bolton, P. and M. Whinston (1993), "Incomplete Contracts, Vertical Integration and Supply Assurance," *Review of Economic Studies*," 60:121-148.

Brusco, S. and F. Panunzi (2000), "Reallocation of Corporate Resources and Managerial Incentives in Internal Capital Markets," mimeo, University of Bocconi, Milan.

Chiu, Y. S. (1998), "Noncooperative Bargaining, Hostages, and Optimal Asset Ownership," *American Economic Review*, 88(4):882-901.

Coase, R. (1988), "Industrial Organization: A Proposal for Research," Chapter 3 in *The Firm, the Market and the Law*. Chicago: The University of Chicago Press.

de Meza, D. and B. Lockwood (1998), "Does Asset Ownership Always Motivate Managers? The Property Rights Theory of the Firm with Alternating-Offers Bargaining," *Quarterly Journal of Economic*s, 113 (2):361- 86.

Elfenbein, D. and J. Lerner (2003), "Ownership and Control Rights in Internet Portal Alliances, 1995-1995," *Rand Journal of Economics*, 34, forthcoming.

Grossman, S. and O. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Econom*y, 94(4):691-719.

Hart, O. (1995), *Firms, Contracts, and Financial Structure* (Oxford: Oxford University Press).

Hart, O. and J. Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Econom*y, 98 (6):1119-1158.

________________ (2000), "On the Design of Hierarchies: Coordination Versus Specialization," mimeo, Harvard University.

Holmstrom, B. (1999a), "The Firm as a Subeconomy," *Journal of Law, Economics and Organization*, 15:74-102.

Holmstrom, B. and J. Roberts (1998), "Boundaries of the Firm Revisited," *Journal of Economic Perspectives,* 12(4):73-94.

Inderst, R. and C. Laux (2000), "Incentives in Internal Capital Markets: Capital Constraints, Competition, and Investment Opportunities," working paper, University College London.

Klein, B., R. G. Crawford and A. Alchian (1978), "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law & Economics*, 21(2):297-326.

Mailath, G. J., V. Nocke, and A. Postlewaite (2002), "The Disincentive Effects of Internalizing Externalities," mimeo, University of Pennsylvania.

Milgrom, P. and J. Roberts (1988), "An Economic Approach to Influence Activities and Organizational Responses," *American Journal of Sociology*, 94 (Supplement), S154-S179.

______________________(1990), "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica,* 58 (6): 1255-1278.

Rajan, R. and L. Zingales (1998), "Power in a Theory of the Firm," *Quarterly Journal of Economics*, 113:387-432.

Rajan, R., H. Servaes and L. Zingales (2000), The Cost of Diversity: The Diversification Discount and Inefficient Investment," *Journal of Finance*, 55(1):35-80.

Rotemberg J. and G. Saloner (1994), "Benefits of Narrow Business Strategies," *American Economic Review*, 84 (4):1330-1349.

_______________________ (2000), "Visionaries, Managers and Strategic Direction," *RAND Journal of Economics*, 31 (4):693-716.

Scharfstein, D. and J. Stein (2000), "The Dark Side of Internal Capital Markets: Divisional Rent-Seeking and Inefficient Investment," *Journal of Finance*, 55:2537-64.

Schoar, A. (2001), "Effects of Corporate Diversification on Productivity," mimeo, Sloan School of Management, MIT.

Shleifer, A. and L. H. Summers (1988), "Breach of Trust in Hostile Takeovers," in *Corporate Take-Overs: Causes and Consequences* (edited by A.J. Auerbach). NBER Project Report. University of Chicago Press.

Stein, J. (1997), "Internal Capital Markets and the Competition for Corporate Resources," *Journal of Finance*, 52:111-133.

______ (2002), "Information Production and Capital Allocation: Decentralized vs Hierarchical Firms," *Journal of Finance*, forthcoming.

Van den Steen, E. (2001), "Organizational Beliefs and Managerial Vision," mimeo, Sloan School of Management, MIT.

Williamson, O. (1975), *Markets and Hierarchies: Analysis and Antitrust Implication* (New York: Free Press).

____________ (1985), *The Economic Institutions of Capitalism* (New York: Free Press).

Woodruff, C., (2002), "Non-Contractible Investments and Vertical Integration in the Mexican Footwear Industry," mimeo, University of California, San Diego.

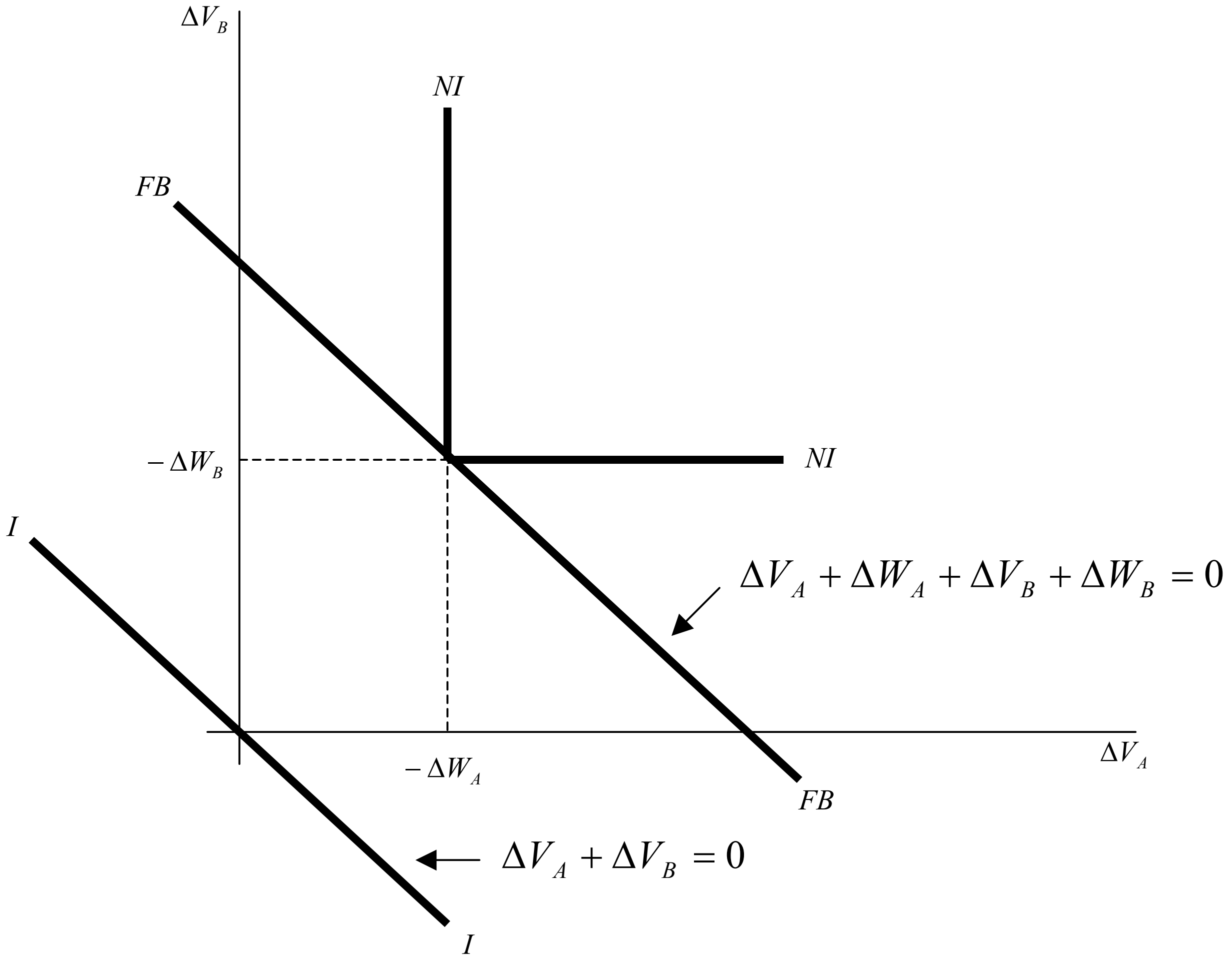
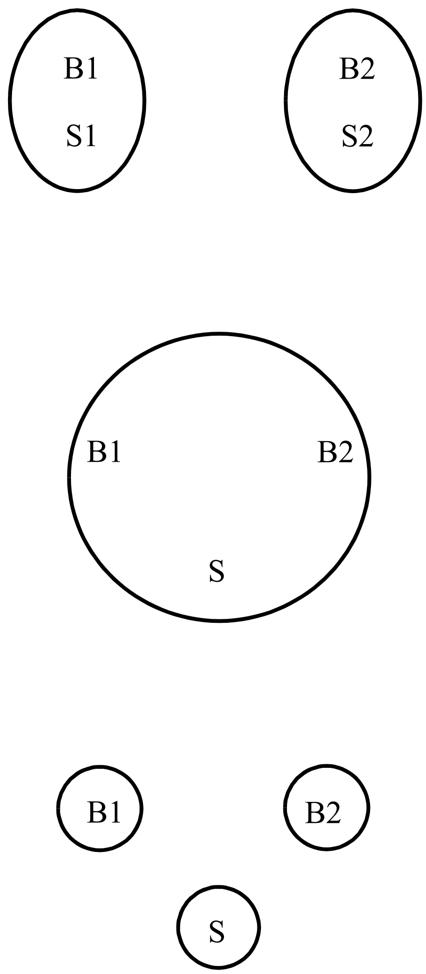

ΔV_B
NI
FB
$-\Delta W_B$
NI
I
$\Delta V_A + \Delta W_A + \Delta V_B + \Delta W_B = 0$
$-\Delta W_A$
ΔV_A
FB
$\Delta V_A + \Delta V_B = 0$
I

Figure 3



(i) Symmetric Vertical Integration

(ii) Complete Integration

(iii) Outsourcing

Figure 5