MECS 570-1: Organizational Economics I

Michael Powell

Boundaries of the Firm, I: Intro, Transaction-Cost Economics, Property Rights Theory

Boundaries of the Firm

The central question in this part of the literature goes back to Ronald Coase (1937): if markets are so great at coordinating productive activity, why is productive activity carried out within firms rather than by self-employed individuals who transact on a spot market? And indeed it is, as Herbert Simon (1991) vividly illustrated:

> A mythical visitor from Mars... approaches Earth from space, equipped with a telescope that reveals social structures. The firms reveal themselves, say, as solid green areas with faint interior contours marking out divisions and departments. Market transactions show as red lines connecting firms, forming a network in the spaces between them. Within firms (and perhaps even between them) the approaching visitor also sees pale blue lines, the lines of authority connecting bosses with various levels of workers... No matter whether our visitor approached the United States or the Soviet Union, urban China or the European Community, the greater part of the space below it would be within the green areas, for almost all inhabitants would be employees, hence inside the firm boundaries. Organizations would be the dominant feature of the landscape. A message sent back home, describing the scene, would speak of "large green areas interconnected by red lines." It would not likely speak of "a network of red lines connecting green spots." ...When our visitor came to know that the green masses were organizations and the red lines connecting them were market transactions, it might be surprised to hear the structure called a market economy. "Wouldn't 'organizational economy' be the more appropriate term?" it might ask.

It is obviously difficult to put actual numbers on the relative importance of trade within and between firms, since, I would venture to say, most transactions within firms are not recorded. From dropping by a colleague's office to ask for help finding a reference, transferring a shaped piece of glass down the assembly line for installation into a mirror, getting an order of fries from the fry cook to deliver to the customer, most economic transactions are difficult even to define as such, let alone track. But we do have some numbers. In what I think is one of the best opening sentences of a job-market paper, Pol Antras provides a lower bound: "Roughly one-third of world trade is intrafirm trade."

Of course, it could conceivably be the case that boundaries don't really matter—that the nature of a particular transaction and the overall volume of transactions is the same whether boundaries are in place or not. And indeed, this would exactly be the case if there were no costs of carrying out transactions: Coase's (1960) eponymous theorem suggests, roughly, that in such a situation, outcomes would be the same no matter how transactions were organized. But clearly this is not the case—in 1997, to pick a random year, the volume of corporate mergers and acquisitions was $1.7 trillion dollars (Holmström and Roberts, 1998). It is implausible that this would be the case if boundaries were irrelevant, as even the associated legal fees have to ring up in the billions of dollars.

And so, in a sense, the premise of the Coase Theorem's contrapositive is clearly true. Therefore, there must be transaction costs. And understanding the nature of these transaction costs will hopefully shed some light on the patterns we see. And as D.H. Robertson also vividly illustrated, there are indeed patterns to what we see. Firms are "islands of conscious power in this ocean of unconscious co-operation like lumps of butter coagulating in a pail of buttermilk." So the question becomes: what transaction costs are important, and how are they important? How, in a sense, can they help make sense out of the pattern of butter and buttermilk?

The field was basically dormant for the next forty years until the early 1970s, largely because "transaction costs" came to represent essentially "a name for the residual"—any

pattern in the data could trivially be attributed to some story about transaction costs. The empirical content of the theory was therefore zero.

Williamson put structure on the theory by identifying specific factors that composed these transaction costs. And importantly, the specific factors he identified had implications about economic objects that at least could, in principle, be contained in a data set. Therefore his causal claims could be, and were, tested. (As a conceptual matter, it is important to note that even if Williamson's causal claims were refuted, this would not invalidate the underlying claim that "transaction costs are important," since as discussed earlier, this more general claim is essentially untestable, because it is impossible to measure, or even conceive of, *all* transaction costs associated with *all* different forms of organization.) The gist of his theory, which we will describe in more detail shortly, is that when contracts are incomplete and parties have disagreements, they may waste resources "haggling" over the appropriate course of action if they transact in a market, whereas if they transact within a firm, these disagreements can be settled by "fiat" by a mediator. Integration is therefore more appealing when haggling costs are higher, which is the case in situations in which contracts are relatively more incomplete and parties disagree more.

But there was a sense in which his theory (and the related work by Klein, Crawford, and Alchian (1978)) was silent on many foundational questions. After all, why does moving the transaction from the market into the firm imply that parties no longer haggle—that is, what is integration? Further, if settling transactions by fiat is more efficient than by haggling, why aren't all transactions carried out within a single firm? Williamson's and others' response was that there are bureaucratic costs ("accounting contrivances," "weakened incentives," and others) associated with putting more transactions within the firm. But surely those costs are also higher when contracts are more incomplete and when there is more disagreement between parties. Put differently, Williamson identified particular costs associated with transacting in the market and other costs associated with transacting within the firm and made assertions about the rates at which these costs vary with the underlying environment. The resulting

empirical implications were consistent with evidence, but the theory still lacked convincing foundations, because it treated these latter costs as essentially exogenous and orthogonal. We will discuss the Transaction-Cost Economics (TCE) approach in the first subsection.

The Property Rights Theory, initiated by Grossman and Hart (1986) and expanded upon in Hart and Moore (1990), proposed a theory which (a) explicitly answered the question of "what is integration?" and (b) treated the costs and benefits of integration symmetrically. Related to the first point is an observation by Alchian and Demsetz that

> It is common to see the firm characterized by the power to settle issues by fiat, by authority, or by disciplinary action superior to that available in the conventional market. This is delusion. The firm does not own all its inputs. It has no power of fiat, no authority, no disciplinary action any different in the slightest degree from ordinary market contracting between any two people. I can "punish" you only by withholding future business or by seeking redress in the courts for any failure to honor our exchange agreement. This is exactly all that any employer can do. He can fire or sue, just as I can fire my grocer by stopping purchases from him or sue him for delivering faulty products.

What, then, is the difference between me "telling my grocer what to do" and me "telling my employee what to do?" In either case, refusal would potentially cause the relationship to break down. The key difference, according to Grossman and Hart's theory, is in what happens after the relationship breaks down. If I stop buying goods from my grocer, I no longer have access to his store and all its associated benefits. He simply loses access to a particular customer. If I stop employing a worker, on the other hand, the worker loses access to all the assets associated with my firm. I simply lose access to that particular worker.

Grossman and Hart's (1986) key insight is that property rights determine who can do what in the event that a relationship breaks down—property rights determine what they refer to as the residual rights of control. And allocating these property rights to one party or another may change their incentives to take actions that affect the value of this particular

relationship. This logic leads to what is often interpreted as Grossman and Hart's main result: property rights (which define whether a particular transaction is carried out "within" a firm or "between" firms) should be allocated to whichever party is responsible for making more important investments in the relationship. We will discuss the Property Rights Theory (PRT) approach in the second subsection.

From a theoretical foundations perspective, Grossman and Hart was a huge step forward—their theory treats the costs of integration and the costs of non-integration symmetrically and systematically analyzes how different factors drive these two costs in a single unified framework. From a conceptual perspective, however, all the action in the theory is related to how organization affects parties' incentives to make relationship-specific investments. As we will see, their theory assumes that conditional on relationship-specific investments, transactions are always carried out efficiently. A manager never wastes time and resources arguing with an employee. An employee never wastes time and resources trying to convince the boss to let him do a different, more desirable task.

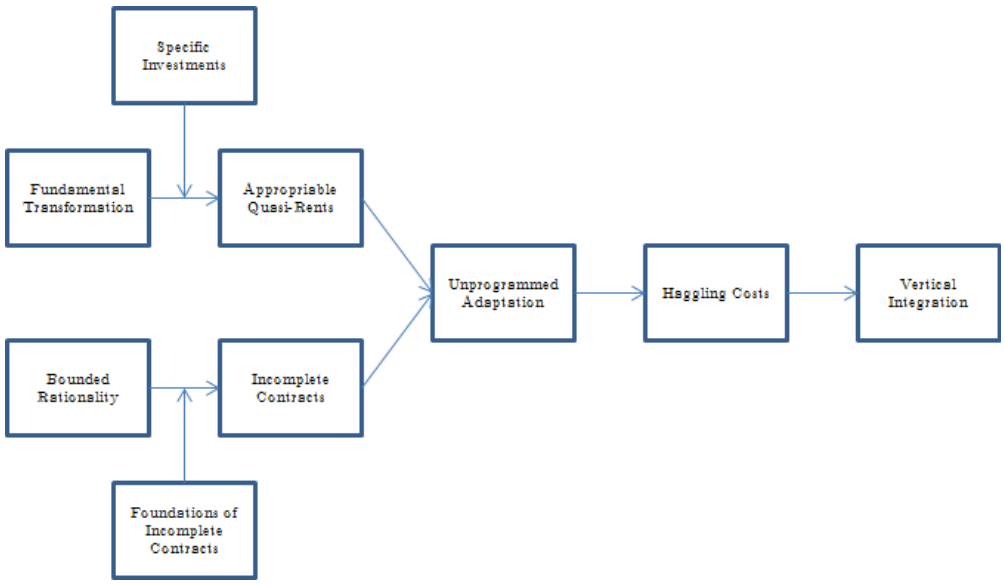
In contrast, in Transaction-Cost Economics, all the action takes place ex post, during the time in which decisions are made. Integration is chosen, precisely because it avoids inefficient haggling costs. We will look at some of the implications of this observation next week in the context of a model in which "haggling costs" are based on influence activities—costly activities aimed at persuading decision makers.

Finally, even the Property Rights Theory does not stand on firm theoretical grounds, since the theory considers only a limited set of institutions the players can put in place to manage their relationship. That is, they focus only on the allocation of control, ignoring the possibility that individuals may write contracts or put in place other types of mechanisms that could potentially do better. In particular, they rule out revelation mechanisms that, in principle, should induce first-best investment. We will address this issue in the sixth subsection.

As a research topic, the theory of the firm can be tricky to navigate. In contrast to many applied-theory topics, the theory of the firm takes a lot of non-standard variables as endogenous. Further, there is often ambiguity about what variables should be taken as endogenous and what methodology should be used, so the "playing field" is not well-specified. But ultimately, I think that developing a stronger understanding of what determines firm boundaries is important, since it simultaneously tells us what the limitations of markets are. I will try to outline some of the "ground rules" that I have been able to discern from spending some time studying these issues.

Transaction-Cost Economics

The literature on the boundaries of the firm introduces many new concepts and even more new terms. So we will spend a little bit of time sorting out the terminology before proceeding. The following figure introduces most of the new terms that we will talk about in this section and in the following sections.



As an overview, the basic argument of the Transaction-Cost Economics approach is the following.

Consider a transaction between an upstream manufacturer and a downstream distributor. Should the distributor **buy** from the manufacturer or should it buy the manufacturer and **make** goods itself? The **fundamental transformation** of ex ante perfect competition among manufacturers for the customer's business to ex post small numbers results *from* **specific investments** and results *in* **appropriable quasi-rents**—ex post rents that parties can potentially fight over, because they are locked in with each other. If the parties have written **incomplete contracts** (the **foundations** for which are that **bounded rationality** limits their ability to foresee all relevant contingencies), then they might find themselves in situations that call for **unprogrammed adaptation**. At this point, they may fight over the appropriate course of action, incurring **haggling costs**. These haggling costs can be reduced or eliminated if either the manufacturer purchases the distributor or the distributor purchases the manufacturer, and they become **vertically integrated**.

Every link in this figure is worth discussing. The **fundamental transformation** is, in my view, the most important economic idea to emerge from this theory. We had known since at least Edgeworth that under bilateral monopoly, many problems were possible (Edgeworth focused on indeterminacy) and perhaps inevitable (e.g., the Myerson-Satterthwaite theorem), but that competition among large numbers of potential trading partners would generally (with some exceptions—perfect complementarities between, say, right-shoe manufacturers and left-shoe manufacturers could persist even if the economy became arbitrarily large) push economies towards efficient allocations. Perfect competition is equivalent to the no-surplus condition (Ostroy, 1980; Makowski and Ostroy, 1995)—under perfect competition, you fully appropriate whatever surplus you generate, so everyone else in the economy as a whole is indifferent toward what you do. As a result, your incentives to maximize your own well-being do not come into conflict with others, so this leads to efficient allocations and nothing worth incurring costs to fight over. The underlying intuition for why large numbers of trading partners leads to efficient allocations is that a buyer can always play a seller and her competitors off each other, and in the limit, the next-best seller is just as

good as the current one (and symmetrically for buyers). Williamson's observation was that after a trading relationship has been initiated, the buyer and the sellers develop ties to each other (**quasi-rents**), so that one's current trading partner is always discretely better than the next best alternative. In other words, the beneficial forces of perfect competition almost never hold.

Of course, if during the ex ante competition phase of the relationship, potential trading partners competed with each other by offering enforceable, complete, long-term trading contracts, then the fact that ex post, parties are locked in to each other would be irrelevant. Parties would compete in the market by offering each other utility streams that they are contractually obligated to fulfill, and perfect competition ex ante would lead to maximized long-term gains from trade.

This is where **incomplete contracts** comes into the picture. Such contracts are impossible to write, because they would require parties to be able to conceive of and enumerate all possible contingencies. Because parties are **boundedly rational**, they will only be able to do so for a subset of the possible states. As a result, ex ante competition will lead parties to agree to incomplete contracts for which the parties will need to fill in the details as they go along. In other words, they will occasionally need to make **unprogrammed adaptations**. As an example, a legacy airline (say, American Airlines) and a regional carrier (say, American Eagle) may agree on a flight schedule for flights connecting two cities. But when weather-related disruptions occur, the ideal way of shifting around staff and equipment depends to a large extent on where both carrier's existing staff and equipment are, and there are simply too many different potential configurations for this. As a result, airlines typically do not contractually specify what will happen in the event that there are weather-related disruptions, and they therefore have to figure it out on the spot.

The need to make unprogrammed adaptations would also not be a problem if the parties could simply agree to bargain efficiently ex post after an event occurs that the parties had not planned for. (And indeed, if there was perfect competition ex post, they would not even

need to bargain ex post.) However, under the TCE view, ex-post bargaining is rarely if ever efficient. The legacy airline will insist that its own staff and equipment are unable to make it, so everything would be better if the regional carrier made concessions, and conversely. Such ex post bargaining inevitably leads either to bad ex post decisions (the carrier with the easier-to-access equipment and staff is not the one who ends up putting it in place) or results in other types of **rent-seeking** costs (time and resources are wasted in the bargaining process). These **haggling costs** could be eliminated if both parties were under the direction of a common headquarters that could issue commands and easily resolve these types of conflicts. This involves setting up a **vertically integrated** organization.

Further, vertically integrated organizations involve **bureaucratic costs**. Reorganization involves setup costs. Incentives are usually low-powered inside organizations. Division managers engage in **accounting contrivances** in order to alter decision making of other divisions or the headquarters. Finally, the contract law that governs decisions made by one division that affect another division differs from the contract law that governs decisions made by one firm that affect another—in essence, the latter types of contracts are enforceable, whereas the former types of contracts are not. This difference in contract law is referred to as **forbearance**.

When would we be more likely to see vertical integration? When the environment surrounding a particular transaction is especially complex, contracts are more likely to be incomplete, or they are likely to be more incomplete. As a result, the need for unprogrammed adaptations and their associated haggling costs will be greater. When parties are more locked in to each other, their ability to access the outside market either to look for alternatives or to use alternatives to discipline their bargaining process is lessened. As a result, there is more to fight over when unprogrammed adaptations are required, and their associated haggling costs will be greater. Additionally, integration involves setup costs, and these setup costs are only worth incurring if the parties expect to interact with each other often. Finally, integration itself involves other bureaucratic costs, and so vertical integration is more appealing if these

costs are low. Put differently, the integration decision involves a trade-off between haggling costs under non-integration and bureaucratic costs under integration. To summarize, the main empirical predictions of the TCE theory are:

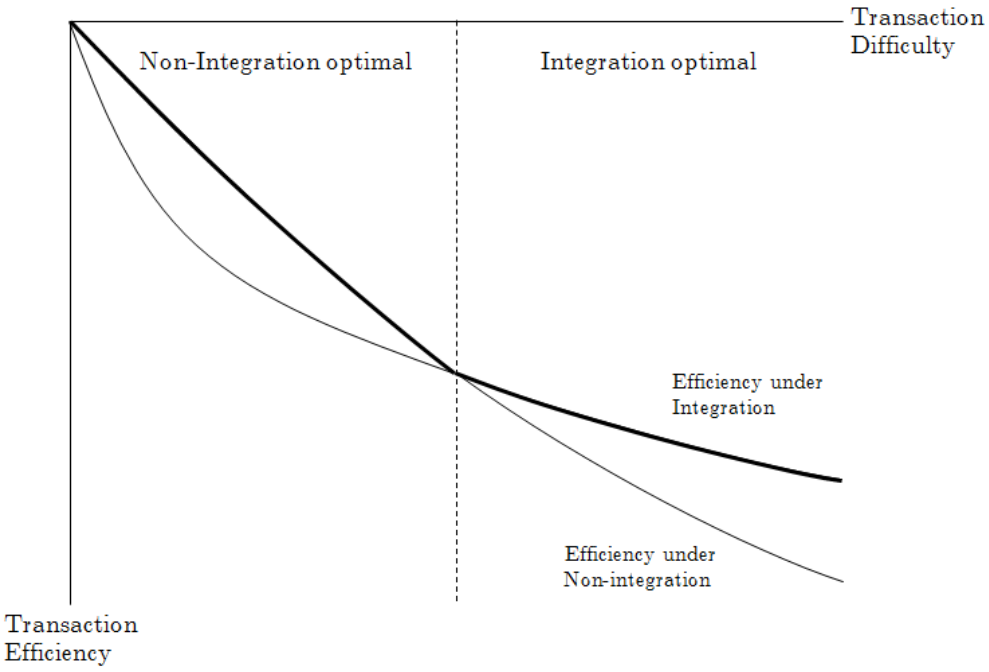
1. Vertical integration is more likely for transactions that are more complicated.
2. Vertical integration is more likely when there is more specificity.
3. Vertical integration is more likely when the players interact more frequently.
4. Vertical integration is more likely when bureaucratic costs are low.

This discussion of TCE has been informal, because the theory itself is informal. There are at least two aspects of this informal argument that can be independently formalized. When unprogrammed adaptation is required, the associated costs can either come from costly haggling (rent-seeking) or from inefficient ex post decision making (adaptation). I will describe two models that each capture one of these sources of ex post inefficiencies.

There are a couple common themes that arise in the analysis of both of these models. The first theme is that when thinking about the make-or-buy or boundary-of-the-firm question, the **appropriate unit of analysis is at the transaction level**. Another theme is that these models consider **private-ordering solutions** rather than solutions imposed on the transacting parties by a third party such as a government—resolving conflict between parties need not involve government intervention. The question is therefore: for a given transaction, what institutions should *the interested parties* put in place to manage this transaction?

Since transactions differ in their characteristics, more difficult transactions will be more difficult no matter how they are organized. As a result, looking at the performance of various transactions and relating that performance to how those transactions are organized could lead one inappropriately to the conclusion that integration is actually bad for performance. Or one could dismiss the agenda, as one prominent economics blogger once did: "I view the Coasian tradition as somewhat of a dead end in industrial organization. Internally, firms

aren't usually more efficient than markets... ."



As the figure above (which Gibbons (2005) colorfully describes as "Coase meets Heckman") shows, this is *exactly* the type of prediction that this class of theories predicts. And whether integrated transactions are less efficient, because integration is bad for transaction efficiency, or because transactions that are more complicated are more likely to involve integration and are likely to be less efficient matters. This difference matters, because these two different views have the opposite implications. Under the TCE view, discouraging firms from engaging in vertical integration (through, for example, strict antitrust policy) will necessarily be bad for firms' internal efficiency. Under the alternative view, strict antitrust policy would serve not only to facilitate product-market competition, it would also *increase* firms' internal efficiency.

Another theme that arises in these models is that there are many differences between transactions carried out between firms and those carried out within firms. An upstream division manager will typically be on a lower-powered incentive scheme than she would be if she were the owner of an independent upstream firm. Transactions within firms are subject

to different legal regimes than transactions between firms. Transactions within firms tend to be characterized by more "bureaucracy" than transactions across firms. There are two ways to look at these bundles of differences. Viewed one way, low-powered incentives and bureaucracy are the baggage associated with integration and therefore a cost of integration. Viewed another way, low-powered incentives and bureaucracy are also optimal choices that complement integration because they help solve other problems that arise under integration.

Finally, I will conclude with a description of one question that I do not think the literature has produced satisfying answers to. First, there are obviously more ways to organize a transaction than "vertical integration" and "non-integration." In particular, the transacting parties could engage in simple spot-market transactions; they could engage in short-run contracting across firm boundaries in which they specify a small number of contingencies; they could engage in long-term contracting across firm boundaries in which perhaps decision rights are contractually reallocated (for example, an upstream firm may have some say over the design specifics for a product that the downstream firm is producing); or one party could buy the other party. The line between integration and non-integration is therefore much blurrier than it seemed at first glance.

Property Rights

Essentially the main result of TCE is the observation that when haggling costs are high under non-integration, then integration is optimal. This result is unsatisfying in at least two senses. First, TCE does not tell us what exactly is the mechanism through which haggling costs are reduced under integration, and second, it does not tell us what the associated costs of integration are, and it therefore does not tell us when we would expect such costs to be high. In principle, in environments in which haggling costs are high under non-integration, then the within-firm equivalent of haggling costs should also be high.

Grossman and Hart (1986) and Hart and Moore (1990) set aside the "make or buy"

question and instead begin with the more fundamental question, "What is a firm?" In some sense, nothing short of an answer to *this* question will consistently provide an answer to the questions that TCE leaves unanswered. Framing the question slightly differently, what do I get if I buy a firm from someone else? The answer is typically that I become the owner of the firm's non-human assets.

Why, though, does it matter who owns non-human assets? If contracts are complete, it does not matter. The parties to a transaction will, ex ante, specify a detailed action plan. One such action plan will be optimal. That action plan will be optimal regardless of who owns the assets that support the transaction, and it will be feasible regardless of who owns the assets. If contracts are incomplete, however, not all contingencies will be specified. The key insight of the PRT is that ownership endows the asset's owner with the right to decide what to do with the assets in these contingencies. That is, ownership confers **residual control rights**. When unprogrammed adaptations become necessary, the party with residual control rights has **power** in the relationship and is protected from expropriation by the other party. That is, control over non-human assets leads to control over human assets, since they provide leverage over the person who lacks the assets. Since she cannot be expropriated, she therefore has incentives to make investments that are specific to the relationship.

Firm boundaries are tantamount to asset ownership, so detailing the costs and benefits of different ownership arrangements provides a complete account of the costs and benefits of different firm-boundary arrangements. Asset ownership, and therefore firm boundaries, determine who possesses power in a relationship, and power determines investment incentives. Under integration, I have all the residual control rights over non-human assets and therefore possess strong investment incentives. Non-integration splits apart residual control rights, and therefore provides me with weaker investment incentives and you with stronger investment incentives. If I own an asset, you do not. Power is scarce and therefore should be allocated optimally.

Methodologically, the PRT makes significant advances over the preceding theory. PRT's

conceptual exercise is to hold technology, preferences, information, and the legal environment constant across prospective governance structures and ask, for a given transaction with given characteristics, whether the transaction is best carried out within a firm or between firms. That is, prior theories associated "make" with some vector $(\alpha_1, \alpha_2, \dots)$ of characteristics and "buy" with some other vector $(\beta_1, \beta_2, \dots)$ of characteristics. "Make" is preferred to "buy" if the vector $(\alpha_1, \alpha_2, \dots)$ is preferred to the vector $(\beta_1, \beta_2, \dots)$. In contrast, PRT focuses on a single aspect: α_1 versus β_1. Further differences may arise between "make" and "buy," but to the extent that they are also choice variables, they will arise optimally rather than passively. We will talk about why this is an important distinction to make when we talk about the influence-cost model in the next section.

Description There is a risk-neutral upstream manager U, a risk-neutral downstream manager D, and two assets A_1 and A_2. Managers U and D make investments e_U and e_D at private cost $c_U(e_U)$ and $c_D(e_D)$. These investments determine the value that each manager receives if trade occurs, $V_U(e_U, e_D)$ and $V_D(e_U, e_D)$. There is a state of the world, $s \in S = S_C \cup S_{NC}$, with $S_C \cap S_{NC} = \emptyset$ and $\Pr[s \in S_{NC}] = \mu$. In state s, the identity of the ideal good to be traded is s—if the managers trade good s, they receive $V_U(e_U, e_D)$ and $V_D(e_U, e_D)$. If the managers trade good $s' \neq s$, they both receive $-\infty$. The managers choose an asset allocation, denoted by g, from a set $G = \{UI, DI, NI, RNI\}$. Under $g = UI$, U owns both assets. Under $g = DI$, D owns both assets. Under $g = NI$, U owns asset A_1 and D owns asset A_2. Under $g = RNI$, D owns asset A_1, and U owns asset A_2. In addition to determining an asset allocation, manager U also offers an incomplete contract $w \in W = \{w : E_U \times E_D \times S_C \to \mathbb{R}\}$ to D. The contract specifies a transfer $w(e_U, e_D, s)$ to be paid from D to U if they trade good $s \in S_C$. If the players want to trade a good $s \in S_{NC}$, they do so in the following way. With probability $\frac{1}{2}$, U makes a take-it-or-leave-it offer $w_U(s)$ to D, specifying trade and a price. With probability $\frac{1}{2}$, D makes a take-it-or-leave-it offer $w_D(s)$ to U specifying trade and a price. If trade does not occur, then manager U receives payoff $v_U(e_U, e_D; g)$ and manager D

receives payoff $v_D(e_U, e_D; g)$, which depends on the asset allocation.

Timing There are five periods:

1. U offers D an asset allocation $g \in G$ and a contract $w \in W$. Both g and w are commonly observed.

2. U and D simultaneously choose investment levels e_U and e_D at private cost $c(e_U)$ and $c(e_D)$. These investment levels are commonly observed by e_U and e_D.

3. The state of the world, $s \in S$ is realized.

4. If $s \in S_C$, D buys good s at price specified by w. If $s \in S_{NC}$, U and D engage in 50-50 take-it-or-leave-it bargaining.

5. Payoffs are realized.

Equilibrium A **subgame-perfect equilibrium** is an asset allocation g^*, a contract w^*, investment strategies $e_U^* : G \times W \to \mathbb{R}_+$ and $e_D^* : G \times W \to \mathbb{R}_+$, and a pair of offer rules $w_U^* : E_D \times E_U \times S_{NC} \to \mathbb{R}$ and $w_D^* : E_D \times E_U \times S_{NC} \to \mathbb{R}$ such that given $e_U^*(g^*, w^*)$ and $e_D^*(g^*, w^*)$, the managers optimally make offers $w_U^*(e_U^*, e_D^*)$ and $w_D^*(e_U^*, e_D^*)$ in states $s \in S_{NC}$; given g^* and w^*, managers optimally choose $e_U^*(g^*, w^*)$ and $e_D^*(g^*, w^*)$; and U optimally offers asset allocation g^* and contract w^*.

Assumptions As always, we will assume $c_U(e_U) = \frac{1}{2}e_U^2$ and $c_D(e_D) = \frac{1}{2}e_D^2$. We will also assume that $\mu = 1$, so that the probability that an ex ante specifiable good is optimal to

trade ex post is zero. We will return to this issue later. Let

$$\begin{aligned} V_U(e_U, e_D) &= f_{UU}e_U + f_{UD}e_D \\ V_D(e_U, e_D) &= f_{DU}e_U + f_{DD}e_D \\ v_U(e_U, e_D; g) &= h^g_{UU}e_U + h^g_{UD}e_D \\ v_D(e_U, e_D; g) &= h^g_{DU}e_U + h^g_{DD}e_D, \end{aligned}$$

and define

$$\begin{aligned} F_U &= f_{UU} + f_{DU} \\ F_D &= f_{UD} + f_{DD}. \end{aligned}$$

Finally, outside options are more sensitive to one's own investments the more assets one owns:

$$\begin{aligned} h^{UI}_{UU} &\geq h^{NI}_{UU} \geq h^{DI}_{UU}, h^{UI}_{UU} \geq h^{RNI}_{UU} \geq h^{DI}_{UU} \\ h^{DI}_{DD} &\geq h^{NI}_{DD} \geq h^{UI}_{DD}, h^{DI}_{DD} \geq h^{RNI}_{DD} \geq h^{UI}_{DD}. \end{aligned}$$

The Program We solve backwards. For all $s \in S_{NC}$, with probability $\frac{1}{2}$, U will offer price $w_U(e_U, e_D)$. D will accept this offer as long as $V_D(e_U, e_D) - w_U(e_U, e_D) \geq v_D(e_U, e_D; g)$. U's offer will ensure that this holds with equality (or else U could increase w_U a bit and increase his profits while still having his offer accepted):

$$\begin{aligned} \pi_U &= V_U(e_U, e_D) + w_U(e_U, e_D) = V_U(e_U, e_D) + V_D(e_U, e_D) - v_D(e_U, e_D; g) \\ \pi_D &= V_D(e_U, e_D) - w_U(e_U, e_D) = v_D(e_U, e_D; g). \end{aligned}$$

Similarly, with probability $\frac{1}{2}$, D will offer price $w_D(e_U, e_D)$. U will accept this offer as long as $V_U(e_U, e_D) + w_D(e_U, e_D) \geq v_U(e_U, e_D; g)$. D's offer will ensure that this holds with

equality (or else D could decrease w_D a bit and increase her profits while still having her offer accepted):

$$\begin{aligned}\pi_U &= V_U(e_U, e_D) + w_D(e_U, e_D) = v_U(e_U, e_D; g) \\ \pi_D &= V_D(e_U, e_D) - w_D(e_U, e_D) = V_U(e_U, e_D) + V_D(e_U, e_D) - v_U(e_U, e_D; g).\end{aligned}$$

In period 2, manager U will conjecture e_D and solve

$$\max_{\hat{e}_U} \frac{1}{2}\left(V_U(\hat{e}_U, e_D) + V_D(\hat{e}_U, e_D) - v_D(\hat{e}_U, e_D; g)\right) + \frac{1}{2}v_U(\hat{e}_U, e_D; g) - c(\hat{e}_U)$$

and manager D will conjecture e_U and solve

$$\max_{\hat{e}_D} \frac{1}{2}v_D(e_U, \hat{e}_D; g) + \frac{1}{2}\left(V_U(e_U, \hat{e}_D) + V_D(e_U, \hat{e}_D) - v_U(e_U, \hat{e}_D; g)\right) - c(\hat{e}_D).$$

Substituting in the functional forms we assumed above, these problems become:

$$\max_{\hat{e}_U} \frac{1}{2}\left(F_U\hat{e}_U + F_D e_D\right) + \frac{1}{2}\left(\left(h^g_{UU} - h^g_{DU}\right)\hat{e}_U + \left(h^g_{UD} - h^g_{DD}\right)e_D\right) - \frac{1}{2}\hat{e}_U^2$$

and

$$\max_{\hat{e}_D} \frac{1}{2}\left(F_U e_U + F_D\hat{e}_D\right) + \frac{1}{2}\left(\left(h^g_{DU} - h^g_{UU}\right)e_U + \left(h^g_{DD} - h^g_{UD}\right)\hat{e}_D\right) - \frac{1}{2}\hat{e}_D^2.$$

These are well-behaved objective functions, and in each one, there are no interactions between the managers' investments, so each manager has a dominant strategy, which we can solve for by taking first-order conditions:

$$\begin{aligned}e^{*g}_U &= \frac{1}{2}F_U + \frac{1}{2}\left(h^g_{UU} - h^g_{DU}\right) \\ e^{*g}_D &= \frac{1}{2}F_D + \frac{1}{2}\left(h^g_{DD} - h^g_{UD}\right)\end{aligned}$$

Each manager's incentives to invest are derived from two sources: (1) the marginal impact

of investment on total surplus and (2) the marginal impact of investment on the "threat-point differential." The latter point is worth expanding on. If U increases his investment, his outside option goes up by h^g_{UU}, which increases the price that D will have to offer him when she makes her take-it-or-leave-it offer, which increases U's ex-post payoff if $h^g_{UU} > 0$. Further, D's outside option goes up by h^g_{DU}, which increases the price that U has to offer D when he makes his take-it-or-leave-it-offer, which decreases U's ex-post payoff if $h^g_{DU} > 0$.

Ex ante, players' equilibrium payoffs are:

$$\begin{aligned}
\Pi^{*g}_U &= \frac{1}{2}\left(F_U e^{*g}_U + F_D e^{*g}_D\right) + \frac{1}{2}\left(\left(h^g_{UU} - h^g_{DU}\right) e^{*g}_U + \left(h^g_{UD} - h^g_{DD}\right) e^{*g}_D\right) - \frac{1}{2}\left(e^{*g}_U\right)^2 \\
\Pi^{*g}_D &= \frac{1}{2}\left(F_U e^{*g}_U + F_D e^{*g}_D\right) + \frac{1}{2}\left(\left(h^g_{DU} - h^g_{UU}\right) e^{*g}_U + \left(h^g_{DD} - h^g_{UD}\right) e^{*g}_D\right) - \frac{1}{2}\left(e^{*g}_D\right)^2.
\end{aligned}$$

If we let $\theta = \left(f_{UU}, f_{UD}, f_{DU}, f_{DD}, \left\{h^g_{UU}, h^g_{UD}, h^g_{DU}, h^g_{DD}\right\}_{g\in G}\right)$ denote the parameters of the model, the Coasian objective for **governance structure** g is:

$$W^g(\theta) = \Pi^{*g}_U + \Pi^{*g}_D = F_U e^*_U + F_D e^*_D - \frac{1}{2}\left(e^{*g}_U\right)^2 - \frac{1}{2}\left(e^{*g}_D\right)^2.$$

The **Coasian Program** that describes the optimal governance structure is then:

$$W^*(\theta) = \max_{g\in G} W^g(\theta).$$

At this level of generality, the model is too rich to provide straightforward insights. In order to make progress, we will introduce the following definitions. If $f_{ij} = h^g_{ij} = 0$ for $i \neq j$, we say that investments are **self-investments**. If $f_{ii} = h^g_{ii} = 0$, we say that investments are **cross-investments**. When investments are self-investments, the following definitions are useful. Assets A_1 and A_2 are **independent** if $h^{UI}_{UU} = h^{NI}_{UU} = h^{RNI}_{UU}$ and $h^{DI}_{DD} = h^{NI}_{DD} = h^{RNI}_{DD}$ (i.e., if owning the second asset does not increase one's marginal incentives to invest beyond the incentives provided by owning a single asset). Assets A_1 and A_2 are **strictly complementary** if either $h^{NI}_{UU} = h^{RNI}_{UU} = h^{DI}_{UU}$ or $h^{NI}_{DD} = h^{RNI}_{DD} = h^{UI}_{DD}$

(i.e., if for one player, owning one asset provides the same incentives to invest as owning no assets). U's **human capital is essential** if $h_{DD}^{DI} = h_{DD}^{UI}$, and D's human capital is essential if $h_{UU}^{UI} = h_{UU}^{DI}$.

With these definitions in hand, we can get a sense for what features of the model drive the optimal governance-structure choice.

PROPOSITION (Hart 1995). If A_1 and A_2 are independent, then NI or RNI is optimal. If A_1 and A_2 are strictly complementary, then DI or UI is optimal. If U's human capital is essential, UI is optimal. If D's human capital is essential, DI is optimal. If both U's and D's human capital is essential, all governance structures are equally good.

These results are straightforward to prove. If A_1 and A_2 are independent, then there is no additional benefit of allocating a second asset to a single party. Dividing up the assets therefore strengthens one party's investment incentives without affecting the other's. If A_1 and A_2 are strictly complementary, then relative to integration, dividing up the assets necessarily weakens one party's investment incentives without increasing the other's, so one form of integration clearly dominates. If U's human capital is essential, then D's investment incentives are independent of which assets he owns, so UI is at least weakly optimal.

The more general results of this framework are that (a) allocating an asset to an individual strengthens that party's incentives to invest, since it increases his bargaining position when unprogrammed adaptation is required, (b) allocating an asset to one individual has an opportunity cost, since it means that it cannot be allocated to the other party. Since we have assumed that investment is always socially valuable, this implies that assets should always be allocated to exactly one party (if joint ownership means that both parties have a veto right). Further, allocating an asset to a particular party is more desirable the more important that party's investment is for joint welfare and the more sensitive his/her investment is to asset ownership. Finally, assets should be co-owned when there are complementarities between them.

While the actual results of the PRT model are sensible and intuitive, there are many

limitations of the analysis. First, as Holmstrom points out in his 1999 JLEO article, "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... Individual ownership of assets does not offer a theory of organizational identities unless one associates individuals with firms." Holmstrom concludes that, "... the boundary question is in my view fundamentally about the distribution of activities: What do firms do rather than what do they own? Understanding asset configurations should not become an end in itself, but rather a means toward understanding activity configurations." That is, by taking payoff functions V_U and V_D as exogenous, the theory is abstracting from what Holmstrom views as the key issue of what a firm really is.

Second, after assets have been allocated and investments made, adaptation is made efficiently. The managers always reach an ex post efficient arrangement in an efficient manner, and all inefficiencies arise ex ante through inadequate incentives to make relationship-specific investments. Williamson (2000) argues that "The most consequential difference between the TCE and GHM setups is that the former holds that maladaptation in the contract execution interval is the principal source of inefficiency, whereas GHM vaporize ex post maladaptation by their assumptions of common knowledge and ex post bargaining." That is, Williamson believes that ex post inefficiencies are the primary sources of inefficiencies that have to be managed by adjusting firm boundaries, while the PRT model focuses solely on ex ante inefficiencies. The two approaches are obviously complementary, but there is an entire dimension of the problem that is being left untouched under this approach.

Finally, in the Coasian Program of the PRT model, the parties are unable to write formal contracts (in the above version of the model, this is true only when $\mu = 1$) and therefore the only instrument they have to motivate relationship-specific investments is the allocation of assets. The implicit assumption underlying the focus on asset ownership is that the characteristics defining what should be traded in which state of the world are difficult to write into a formal contract in a way that a third-party enforcer can unambiguously

enforce. State-contingent trade is therefore unverifiable, so contracts written directly or indirectly on relationship-specific investments are infeasible. However, PRT assumes that relationship-specific investments, and therefore the value of different ex post trades, are commonly observable to U and D. Further, U and D can correctly anticipate the payoff consequences of different asset allocations and different levels of investment. Under the assumptions that relationship-specific investments are commonly observable and that players can foresee the payoff consequences of their actions, Maskin and Tirole (1999) show that the players should always be able to construct a mechanism in which they truthfully reveal the payoffs they would receive to a third-party enforcer. If the parties are able to write a contract on these announcements, then they should indirectly be able to write a contract on ex ante investments. This debate over the "foundations of incomplete contracting" mostly played out over the mid-to-late 1990s, but it has attracted some recent attention. We will discuss it in more detail later.