
MASSACHUSETTS
LIBRARIES
INSTITUTE of TECHNOLOGY

Massachusetts Institute of Technology
Department of Economics
Working Paper Series

Four Formal(izable) Theories of the Firm?

Robert Gibbons

Working Paper 04-34

September 16, 2004

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Four Formal(izable) Theories of the Firm?

Robert Gibbons*
MIT and NBER
rgibbons@mit.edu

September 16, 2004

Abstract

In this essay, I define and compare elemental versions of four theories of the firm. These elemental theories are distilled from important contributions by Hart, Holmstrom, Klein, Williamson, and others. Although these contributions have been widely cited and much discussed, I have found it difficult to understand the commonalities, distinctions, and potential combinations of these seemingly familiar contributions. In this essay, therefore, I attempt to clarify these issues, in three steps: I begin with informal summaries of the theories, then turn to simple but formal statements of each elemental theory, and finally nest the four elemental theories in an integrative framework.

* I thank Matthew Bidwell and Marek Pycia for exemplary research assistance, George Baker, Oliver Hart, Paul Joskow, Kevin J. Murphy, and Michael Whinston for helpful comments, Pierre Azoulay, Jacques Crémer, Luis Garicano, Bengt Holmstrom, Ben Klein, Claude Menard, John Roberts, Eric Van den Steen, Michael Waldman, and Oliver Williamson for detailed suggestions, and Jonathan Levin, Scott Masten, Stéphane Saussier, and Steven Tadelis for sustained dialogues on the frailties of this essay. I also thank seminar audiences at ATOM (Paris I), Berkeley, Chicago, Harvard-MIT, ISNIE, Stanford, and USC for helpful comments and MIT's Program on Innovation in Markets and Organizations, Harvard Business School, and the NSF (SBR-9809107) for financial support. This essay will evolve into two chapters in my forthcoming book, *Organizational Economics* (Princeton University Press, 2006), so comments are still very welcome.

Four Formal(izable) Theories of the Firm?

by

Robert Gibbons

After halting beginnings, the theory of the firm has become a big business. Coase (1937) posed the theory's defining question: which transactions are more efficiently conducted in a firm than in a market? But then the field lay fallow for several decades. Since the 1970s, however, the theory of the firm has become one of the most fertile fields in the profession.

In this essay, I define and compare elemental versions of four theories of the firm: (1) a "rent seeking" theory, which can be discerned in informal theoretical arguments by Williamson (1971, 1979, 1985) and Klein, Crawford, and Alchian (1978) and in early empirical work by Monteverde and Teece (1982), Anderson and Schmittlein (1984), Masten (1984), and Joskow (1985); (2) a "property rights" theory, which can be discerned in formal models by Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995); (3) an "incentive system" theory, which can be discerned in formal models by Holmstrom and Milgrom (1991, 1994), Holmstrom and Tirole (1991), and Holmstrom (1999); and (4) an "adaptation" theory, which can be discerned in informal theoretical arguments by Simon (1951), Williamson (1971, 1973, 1975, 1991), Klein and Murphy (1988, 1997), and Klein (1996, 2000a).

Although these seemingly familiar contributions have been widely cited and much discussed, I have found it difficult to understand their commonalities, distinctions, and potential combinations. Furthermore, it seems unlikely that I am alone. For example, one still sometimes hears the claim that "Grossman and Hart (1986) formalized Williamson (1979)." Indeed, I have heard this claim with two opposite spins: "Grossman-Hart *merely* formalized Williamson," and "*Finally*, someone formalized Williamson." But regardless of spin, this claim is just plain wrong: as will become clear below, Grossman-Hart and Williamson offer different theories. (In fact, in the elemental versions I present, these theories are essentially orthogonal.)

I proceed in three steps: I begin with informal statements of the four theories, then turn to simple but formal statements of each elemental theory, and finally nest the four theories in an integrative framework. These formal statements are "elemental" in the

dictionary's sense of being "reduced to stark simplicity," so they may sacrifice much of the richness and insight of the original contributions. But one virtue of these formal statements is that they can be nested in an integrative framework, and many existing and prospective theoretical and empirical projects can then be seen as combining different elements from this integrative framework.

There is some value in clarifying these four theories as abstract models, but I think greater value stems from these theories' wide-ranging current and potential applications. Of course, there are direct applications: to vertical integration, horizontal integration (of diverse business units, as in "corporate strategy"), contracts between firms, and hybrid governance structures (joint ventures, networks, and so on). But there are also indirect applications: one sees these theories at work not only in industrial organization and organizational economics, but also in corporate finance, development economics, economic history, international trade, labor economics, political economy, and beyond. Consequently, I think it is important to get these core models straight.

To conclude this Introduction, let me emphasize four caveats, each of which I discuss further in the Conclusion. The first caveat concerns scope: this essay focuses on "the theory of the firm" (*i.e.*, Coase's make-or-buy problem), with scant attention paid to related topics in organizational economics (such as structures and processes inside organizations). Thus, a more accurate title might have been "Four Formal(izable) Theories of the Boundary of the Firm," but the make-or-buy problem has come to be called the theory of the firm (*e.g.*, in Hart (1995)), so I continue this usage, but I think it is important not to lose sight of the broader usage intended by, say, Cyert and March's (1963) *A Behavioral Theory of the Firm*, in which "theory of the firm" means descriptive and prescriptive models of firms' decision-making processes. See Holmstrom and Tirole (1989) for an early survey that begins with the make-or-buy problem and then discusses capital structure, hierarchies, and other aspects of the broadly defined "Theory of the Firm."

The second caveat concerns coverage: this essay is not a survey; instead, I focus on just four theories, almost entirely ignoring a wide variety of others – formal, informal, and gestating. Holmstrom and Roberts (1998) fill some of this gap by offering wide-ranging discussions of empirical and theoretical considerations that the rent-seeking and property-rights theories seem not to address. Continuing in this spirit, let me emphasize three considerations that seem outside all four of the theories I discuss here: resources (*e.g.*, Penrose, 1959; Wernerfelt, 1984), routines (*e.g.*, Nelson and Winter, 1982;

Henderson and Clark, 1990), and knowledge (*e.g.*, Kogut and Zander, 1992; Nonaka and Takeuchi, 1995). As far as I know, the literatures on resources, routines, and knowledge are (a) largely about internal structures and processes and (b) largely informal, but I think these issues have mouth-watering potential implications for the make-or-buy problem, so I expect them to play key roles in future formal theories of the firm.

The third caveat concerns methodology: this essay discusses theory, essentially ignoring evidence and testing. Whinston (2003) takes a complementary approach, re-examining some of the classic empirical work on the make-or-buy problem to see what light it sheds on subsequent theoretical developments. In particular, Whinston re-examines evidence commonly interpreted as supporting the rent-seeking theory and asks how it squares with the property-rights theory. It would be very valuable to apply such a meta-analytic approach even more broadly, by re-examining a larger set of existing empirical contributions to see what light they shed on all four of the theories I describe.

The fourth (and perhaps most important) caveat concerns style: this essay presents stick-figure introductions, not full-bodied statements of the original theories or their subsequent elaborations. With respect to the existing formal theories (property rights and incentive systems), this caveat warns that I will attempt to distill these theories to their essences – an act involving judgment (and hence some chance of error, or at least disagreement). But for these two theories, I am comforted by the fact that the stick-figure introductions I offer are special cases of the respective formal theories. With respect to the originally informal theories (rent-seeking and adaptation), however, this final caveat has even greater significance: because the arguments by Klein and Williamson are informal, the stick-figure introductions I offer may bear little relation to the arguments that these authors intended. If so, I will take some solace from the fact that the formal models presented here will then apparently be new theories of the firm. But, as I explain in the Conclusion, my main reaction to any remaining gaps between the formal models presented here and the detailed descriptions and informal theories developed by Klein and Williamson is perhaps a surprising one: gratitude. Without three decades of work by these two tireless contributors, I expect that we would have fewer and narrower formal theories, and I am sure that we would be less well positioned to push our current and prospective formal theories towards a thorough understanding of organization and governance.

1. Informal Statements

Terms such as "transaction costs," "property rights," "hold-up," and the like have been given different (explicit and implicit) meanings in the literature. Consequently, there may be some confusion about what different theories actually say. By presenting formal statements of four elemental theories, I hope to contribute precise terminology and distinctions to the literature. But some readers may prefer words to formal statements, so in this section I offer informal statements of the four theories described below. I should caution, however, that while these informal statements are in words rather than math, they are still rather abstract. Therefore, for some readers, especially those with little or no prior exposure to these theories, the informal statements in this section and the formal statements that follow may be best seen as complementary.

Before diving into the distinctions among these theories, it may help to recall what they all should have in common: a theory of the firm must define "integration" (*i.e.*, whether a given transaction is within one firm or between two) and show why it matters (*i.e.*, what tradeoff exists between integration and non-integration, so that the theory predicts integration for some transactions and non-integration for others). For example, the implicit definition of integration in the industrial organization literature is the unification of control rights. This definition of integration explains why an integrated supply chain avoids double marginalization (vertical integration) and why the price level under monopoly differs from the price level under duopoly (horizontal integration). But while the IO literature defines integration, it does not describe a tradeoff that integration creates; that is, in and of itself, the definition of integration as the unification of control rights offers no downside to integration, so the implicit prediction is that we should observe one gigantic firm. Most of the IO literature seems content to ignore this issue, perhaps implicitly appealing to antitrust constraints as a limit to integration.

The early (informal) theories of the firm reviewed here also adopted the definition of integration as the unification of control rights; for example, see Williamson's (1971: 114) discussion of how integration allows decisions to be made by "fiat." Relative to the IO literature, however, these early theories of the firm clearly recognized the crucial role of contract imperfections in making integration matter; for example, Williamson (1971: 113) argues that "only when the need to make unprogrammed adaptations is introduced does the market versus internal organization issue become engaging." Furthermore, these early theories explicitly recognized the need to provide a downside to integration, but largely left this task for future research; for example, Williamson (1971: 113) notes that

"a complete treatment of vertical integration requires that the limits as well as the powers of internal organization be assessed."

As the earliest formal theory of the firm reviewed here, Grossman and Hart's (1986) model was broadly consistent with two aspects emphasized in much of the earlier informal theory: the definition of integration as the unification of control and the focus on contract imperfections as a necessary condition for integration to matter. (More precisely, the first step in the Grossman-Hart argument was to posit a world with incomplete contracts. In such a world, ownership conveys "residual rights of control" – that is, all the decision rights not specified in a contract.) Where Grossman and Hart really made progress, however, at least in my view, was in delivering a unified account of the costs and benefits of integration. That is, holding the economic environment fixed, they analyzed alternative governance structures and compared the results. (To mix sporting metaphors, one might say that Grossman and Hart ran the Coasean horserace on a level playing field.) In the discussion of the property-rights theory of the firm below, I give the specific Grossman-Hart arguments for why integration matters in the economic environment they postulated. But in all the theories discussed below (not just the property-rights theory), I adopt the Grossman-Hart goal of giving a unified account of the costs and benefits of integration.

1A. Rent Seeking

In the rent-seeking theory of the firm – Williamson (1971, 1979, 1985), Klein, Crawford, and Alchian (1978) – integration can stop socially destructive haggling over "appropriable quasi-rents" (hereafter AQRs).[1] For example, Williamson (1971: 114-5) argues that "fiat is frequently a *more efficient* way to settle minor conflicts ... than is haggling" (emphasis added). The key idea is that, in some circumstances, non-integration cannot avoid inefficient haggling because, while "jointly (and socially) unproductive, it constitutes a source of private pecuniary gain," so integration (with dispute-resolution by fiat) will be more efficient.

[1] Let me reiterate the final caveat from the Introduction. By listing Klein, Crawford, and Alchian (1978) and Williamson (1971, 1979, 1985) as key sources for the rent-seeking theory, I mean that (1) the key idea of this theory (that integration can stop socially destructive haggling over AQRs) is a prominent feature of these papers, and that (2) I do not know of earlier work that prominently features this key idea. I do *not* mean that this idea is the only idea in these papers. For example, Klein-Crawford-Alchian and Williamson also make arguments similar to the property-rights theory described below.

As I discuss in Section 3, the rent-seeking theory of the firm seems never to have been fully formalized (in the sense of producing a unified account of the costs and benefits of integration), but its classic case is the General Motors-Fisher Body relationship, as described by Klein-Crawford-Alchian and further analyzed by Klein (1988, 2000b, 2004). In this case, the formal contract between GM and Fisher is said to have encouraged Fisher to take socially inefficient actions (*e.g.*, regarding plant location), so as to increase Fisher's profit, at disproportionate expense to GM. To stop this hold-up, GM eventually acquired Fisher.[2]

Although the rent-seeking theory was not fully formalized, this did not prevent empirical work motivated by this theory, beginning with now-classic contributions such as Monteverde and Teece (1982), Anderson and Schmittlein (1984), Masten (1984), and Joskow (1985). The unifying theme in this empirical work follows the GM-Fisher story: larger AQRs make integration more likely, presumably because larger AQRs make socially destructive haggling either more likely or more costly or both. For purposes of comparison with the property-rights theory described next, it is useful to note that this prediction – that larger AQRs make integration more likely – does not require knowing the source of the AQRs. In particular, in the rent-seeking theory, either contractible or non-contractible specific investments could create the AQRs that inspire inefficient haggling. For example, imagine that GM and Fisher Body wrote a perfect contract specifying exactly where a plant should be located and how it should be designed and built: once the plant is built, precisely to specifications, there are appropriable quasi-rents to be had. In contrast, we will see that in the property-rights theory, it is crucial that the specific investments be non-contractible.

One feature of this (and any other) informal theory is that its assumptions are not entirely clear. For example, the rent-seeking theory explicitly assumes that integration can stop the haggling induced by AQRs, but this explicit assumption requires an implicit focus on certain kinds of haggling. Specifically, if the haggling were accomplished by manipulation of alienable (say, physical) capital, then integration could remove the relevant control rights from the haggler, but if the haggling were accomplished by

[2] More detailed investigation of the GM-Fisher case has produced many additional facts, some of which seem counter to the simple story told above; see Coase (2000), Freeland (2000), Casadesus-Masanell and Spulber (2000), and Helper, MacDuffie, and Sabel (2000). Regardless of the facts of the GM-Fisher case, however, it is clear that the simple story told above inspired Klein to sketch an informal theory that has broad potential applicability. Since the purpose of this paper is to articulate elemental formal theories, it is irrelevant to my present purposes whether the case that inspired Klein was fact, myth, or some combination.

manipulation of inalienable (say, human) capital, then integration could not stop rent-seeking. More generally, the most that integration can do is to unify the alienable control rights; any inalienable control rights are staying put, by definition. Thus, the distinctive point in (this telling of) the rent-seeking theory of the firm is that *ownership can stop haggling that is undertaken via alienable instruments.*

But we observe many hold-ups between firms that do not result in integration. To explain these observations, the rent-seeking theory has two options: (1) assert that these hold-ups utilized inalienable instruments (so that the observed hold-ups are unavoidable), or (2) enrich the theory to include a downside of integration (so that the observed hold-ups are a lesser evil than integration would have been). As I have so far told it, however, the rent-seeking theory says nothing about what life was like as the Fisher division of General Motors, and hence gives no insight into whether integration could ever be the greater of two evils. As a result, the prediction I stated above is flawed: so far, we can conclude that larger AQRs make non-integration more costly, but we cannot draw an inference about the likelihood of integration until we say something about the costs of integration. I return to this issue in Section 3.

1B. Property Rights

As noted above, one of the key contributions of Grossman and Hart (1986) was that it gave a unified account of the costs and benefits of integration – a challenge that all theories of the firm must confront but that no formal theory had previously overcome. Viewed more narrowly, as one particular theory of the firm, however, the property-rights theory of the firm – Grossman and Hart (1986), Hart and Moore (1990), Hart (1995) – can be seen as the inverse of the rent-seeking theory. Specifically, where the rent-seeking theory envisions socially destructive haggling ex post, the property-rights theory assumes efficient bargaining, and where the rent-seeking theory is consistent with contractible specific investments ex ante, the property-rights theory requires non-contractible specific investments. These distinctions should already make it clear that the property-rights theory in no sense formalizes the rent-seeking theory (*i.e.*, Grossman-Hart did not formalize Williamson, as Holmstrom and Roberts (1998) and Whinston (2003) also emphasize).

In the property-rights theory, efficient bargaining causes the parties to share the surplus from their specific investments. Each party's surplus share determines that party's investment incentive. In turn, each party's asset ownership determines that party's

surplus share. Under certain separability assumptions, owning more assets guarantees a bigger surplus share and so creates a stronger investment incentive. Thus, under these assumptions, if it is important to maximize one party's investment then that party should own all the assets, whereas if the parties' investment incentives are both important then dividing the assets between the parties is efficient. In summary, in the property-rights theory, the integration decision determines ex ante investments and hence total surplus, whereas in the rent-seeking theory, the integration decision determines ex post haggling and hence total surplus.

One insight from (this telling of) the property-rights theory is that *using a formal instrument to stop one hold-up problem typically creates another hold-up problem.* For example, having GM acquire Fisher may stop Fisher's hold-up of GM, but the acquisition also creates the reverse possibility: GM could now hold up its Fisher division, extracting rents created by specific investments that Fisher might undertake. A second insight (a corollary of the first, but cast more concretely) is that *the cost of control is the loss of initiative.* For example, after integration, if GM can now hold up its Fisher division, the prospect of this reverse hold-up may reduce the Fisher division's incentive to invest. In this way, the property-rights theory answers a key question raised in response to the rent-seeking theory: one downside of integration is reduced initiative (for the acquired party).[3]

A related difference between the rent-seeking and property-rights theories concerns internal organization. Whereas the rent-seeking theory (as I have so far told it) is silent about internal organization, an important feature of the property-rights theory is not only that it defines and evaluates life under integration, but also that it does so for the same environment for which it defined and evaluated life under non-integration. Without this feature, the property-rights theory could not provide a unified account of the costs and benefits of integration (*i.e.*, run the Coasean horserace on a level playing field). And yet, the property-rights theory does not paint an attractive picture of life under integration: this is a theory of solo entrepreneurs (single actors who own entire asset combinations) and drone employees (who own nothing and hence, in this model, face no incentives and

[3] Although it is commonplace to discuss hold-up between firms, it may be less familiar to consider hold-up within a firm, such as we are now considering between GM and its Fisher division, so consider the following example. The large pharmaceutical firm Johnson & Johnson consists of hundreds of local operating companies (LOCs), each with its own tightly circumscribed agenda set by headquarters. One of these LOCs handles Tylenol, but it is up to headquarters to determine whether this LOC or another handles closely related products, such as Tylenol with codeine. Employees of the Tylenol LOC could feel held-up if they worked hard to develop Tylenol with codeine, only to be told that another LOC would handle that new product. See Aghion and Tirole (1997) and Baker, Gibbons, and Murphy (1999) for more on such hold-ups within firms.

so do nothing). Where are delegation and internal incentives, to enrich the conception of the drone employees? Furthermore, where are managers (*i.e.*, non-owners who manage someone)?

1C. Incentive Systems

In its simplest form, the incentive-system theory of the firm – Holmstrom and Milgrom (1991, 1994), Holmstrom and Tirole (1991), Holmstrom (1999) – can be seen as an accidental theory of the firm. Instead of focusing on the make-or-buy problem that motivated the rent-seeking and property-rights theories, this work focuses on an incentive problem between a principal and an agent. As a result of this agency-theoretic focus, however, the incentive-system theory of the firm has the virtue of analyzing internal incentives, avoiding the drone employees of the property-rights theory.

In the incentive-system theory, there are many tasks and many instruments in an agency problem, where asset ownership is merely one of the instruments. For example, suppose that there are two tasks, where the agent's output is a linear combination of these two tasks, but the agent's measured performance is a different linear combination. In addition, suppose that there is an asset used in the production process, where the value of the asset after production occurs is yet another linear combination, but the value of this asset is not contractible (*i.e.*, whoever owns the asset receives its value). We then have two ways to structure the agency problem: (1) if the agent does not own the asset (and so is an "employee"), all her incentives come from being paid on measured performance; alternatively, (2) if the agent does own the asset (and so is an "independent contractor"), she receives not only a payment based on measured performance but also the asset's value after production occurs, so she has two sources of incentives.

For fixed parameters, the optimal incentive contract is different for an employee than for an independent contractor, because the employee is not distracted by the incentives to increase the asset's value that a contractor faces. In particular, for a broad class of parameters, the optimal contract for an employee provides weaker incentives than does the optimal contract for a contractor. In this sense, the incentive-system theory provides a potential explanation for the widespread (but largely anecdotal) view that incentives offered to employees in firms are low-powered, relative to the high-powered incentives offered to independent contractors in markets. In this explanation, firms and markets have access to the same feasible set of incentive contracts, but firms choose low-

powered incentives because the agent responds to a given contract differently as an employee than she would as a contractor.

But the efficient way to structure the agency problem will optimize not just the incentive contract but also asset ownership. For some parameters, employment is optimal; for others, contracting. More specifically, because the asset's value is not contractible, having the agent own the asset can provide incentives that cannot be replicated via contract. But providing the agent with these incentives to increase asset value may or may not help the principal to influence the agent's incentives via contract. If the incentives from asset ownership hurt the principal's efforts to create incentives via contract, then the principal should own the asset (*i.e.*, integration is efficient). This line of reasoning provides a second (complementary) potential explanation for low-powered incentives in firms: parties choose to transact in firms precisely when transacting in a market would produce misdirected incentives. Where the first explanation conducted a thought experiment with fixed parameters, this explanation compares observed governance structures across different parameters.

In sum, the distinctive point in (this telling of) the incentive-system theory is that *asset ownership can be an instrument in a multi-task incentive problem.* Asset ownership has both direct effects (incentives from asset value) and indirect effects (changes in the optimal incentive contract). Joint optimization over asset ownership and contract parameters illustrates the system approach to incentive problems.

Like the property-rights theory, the incentive-system theory of the firm provides a unified account of the costs and benefits of integration. Furthermore, the incentive-system theory remedies one shortcoming of the property-rights theory: employees now face incentives and so no longer act like drones. But most employees are not governed by formal incentive contracts (MacLeod and Parent, 1999). And more importantly, the elemental incentive-system theory omits one of the central and appealing aspects of the rent-seeking and property-rights theories: control. That is, in the elemental incentive-system theory, whether the agent owns the asset affects the agent's payoff function, but not the agent's action space.

1D. Adaptation

An adaptation theory of the firm – Simon (1951), Williamson (1971, 1973, 1975, 1991), Klein and Murphy (1988, 1997), and Klein (1996, 2000a) – asks whether

integration or non-integration better facilitates "adaptive, sequential decision-making" (Williamson, 1975: 40) in environments where uncertainty is resolved over time. The key theoretical challenge in developing such a theory is to define an environment in which neither contracts ex ante nor renegotiation ex post can induce first-best adaptation after uncertainty is resolved, so that the second-best solution may be to concentrate authority in the hands of a "boss," who then makes (potentially self-interested) decisions after uncertainty is resolved. This emphasis on the boss's authority places the adaptation theory together with the rent-seeking theory in making control the central issue in the theory (whereas the incentive-system theory ignores control in favor of incentives and the property-rights theory blends the two).

We have already seen that Williamson's 1971 paper contained key ideas for the rent-seeking theory of the firm. But this remarkable paper also hinted at an adaptation theory of the firm, arguing that "only when the need to make unprogrammed adaptations is introduced does the market versus internal organization issue become engaging" (p. 113). Williamson then developed this idea much further in his 1975 book. Specifically, Chapter 4 of the 1975 book used Simon's (1951) "Formal Theory of the Employment Relationship" to explain why many labor transactions are more efficiently conducted in a firm instead of in a market.

In Simon's model (which is cast as a theory of employment, rather than a theory of the firm), two parties choose between (a) negotiating a decision before uncertainty is resolved or (b) allocating authority to one party (the "boss"), who can then make a self-interested decision after uncertainty is resolved. Simon calls the latter an employment contract. Under such a contract, the subordinate faces a tradeoff between flexibility and exploitation: she can sacrifice flexibility by locking in a decision now, or she can risk exploitation by allowing the boss to decide later. Simon provides plausible conditions (roughly, that the parties' payoffs depend importantly on tailoring the decision to the state, and that the parties' preferences regarding such tailoring are not too divergent) under which it is optimal for the parties to choose the employment contract.

Having made Simon's model the centerpiece of Chapter 4, Williamson then makes an explicitly parallel case for intermediate products in Chapter 5: "The argument here really parallels that of Chapter 4 in most essential respects" (Williamson 1975: 99). To me, this quotation is a striking (if apparently unremarked) aspect of the field's intellectual history. Simon's 1951 paper is not even cited in Williamson's landmark 1971 or 1979 papers, which are among the seminal works on the rent-seeking theory of the firm

described above. But Simon's paper is absolutely central to Williamson's 1975 book – first as the explicit core of Chapter 4 on labor transactions, and then (by self-proclaimed parallel) as the implicit core of Chapter 5 on make-or-buy.

One way to explain why I find it so interesting that Simon's paper received this central role in Williamson (1975) is to note that Simon says absolutely nothing about specific investments. Thus, I do not conclude from this close textual analysis that Williamson has been inconsistent or confused or wrong; rather, I conclude that his collected works suggest *two* theories of the firm – rent-seeking and adaptation. Much of the literature has focused on rent-seeking, often with AQRs created by specific investments, and sometimes without any mention of adaptation. Williamson himself typically emphasizes both asset specificity and adaptation – probably reflecting the view that both will be important if a full-blown theory of the firm is to be realistic, but possibly reflecting the view that both are necessary if an elemental theory of the firm is to be coherent. As we will see in Section 2B, however, *there can be a coherent elemental theory of the firm without specific investments.*

In Chapter 4 of his 1975 book, Williamson argues not just that firms can facilitate adaptation, but more specifically that firms can facilitate *relational* adaptation – where today's decisions are taken after today's uncertainty is resolved, and then the parties decide whether to repeat this process tomorrow, so the shadow of tomorrow can influence today's behavior. He briefly considers whether markets could facilitate relational adaptation (pp. 107-8), but downplays this possibility.[4] Fortunately, Klein has spent two decades developing the complementary story – see Klein and Leffler (1981), Klein and Murphy (1988, 1997), and especially Klein (1996, 2000a). Like Williamson, Klein begins by emphasizing that contracts between firms are incomplete, but Klein then emphasizes that successful transactions *between* firms often achieve adaptation by using relationships. Ironically, while Klein emphasizes the part of the story that Williamson downplayed (relational contracting between firms), he downplays the part that Williamson emphasized (relational contracting within firms). As I describe in the

[4] In subsequent work, Williamson gives greater credence to the possibility of relational adaptation between firms. For example, Williamson (1979; 1985, Chapter 3) adopts Macneil's distinctions among classical, neoclassical and relational contacting and then prescribes non-integrated relational contracting for some transaction types and integrated relational contracting for others. So I do not mean that Williamson has ignored relational contracting between firms, but I do think he has not given this issue the attention it deserves. For example, by 1991, Williamson had abandoned the relational-contracting part of Macneil's distinctions, arguing instead that market, hybrid, and hierarchy are supported by classical, neoclassical, and forbearance contract law, respectively. I continue this discussion in the Conclusion.

Conclusion however, Williamson's and Klein's approaches can be combined into a single model, in order to determine whether integration or non-integration facilitates superior relational adaptation. The key point in the resulting relational-adaptation theory of the firm is then that *the integration decision is chosen to facilitate the parties' relationship.*

2. Three Formal Theories of the Firm

In this section, I offer stick-figure renditions of three of the four theories described above: property-rights, incentive-system, and adaptation. The property-rights and incentive-system theories were formalized years ago, so it is straightforward to present simple versions of these theories, and it now is also possible to formalize the adaptation theory by borrowing a recently developed modeling approach, but I defer discussion of the rent-seeking theory until Section 3 because I find it tricky to formalize.

For expositional convenience, I discuss the theories in a new order. I begin with the two simplest: incentive-system and adaptation. We will see that, in the incentive-system theory, the asset is a pure payoff right (with no effect on any party's feasible actions), whereas in the adaptation theory, the asset is a pure decision right (with no effect on any party's payoff function). I then describe the property-rights theory, which is more complex to describe, in part because it involves aspects of both the payoff-right formulation and the decision-right formulation.

In all three models, I focus on the simplest possible make-or-buy setting: there are two parties and one asset, so the key question is who should own the asset. Strictly speaking, therefore, these are not models of "integration," since that term presumably requires two or more assets that could be owned together; instead, one could call these models of "ownership." To relate these one-asset models to the discussions of "integration" above, one could imagine that there is an unmodeled asset that is owned by party 1, so that integration occurs if party 1 owns the modeled asset, but non-integration occurs if party 2 owns the modeled asset. Alternatively, one could interpret the one-asset models literally (namely, as determining who should own the single asset in question), understanding that the natural extension to two assets would allow an investigation of integration.

In addition to taking some license with the word "integration," I take perhaps even more license with the words "asset" and "ownership," as follows. In the incentive-system

theory, the key idea is that owning an asset gives the owner title to the asset's value; in this context, "asset" and "ownership" seem like excellent words. In the adaptation theory, however, the key idea is that owning an asset allows the owner to dictate how the asset is used; in this context, the asset might simply be a "decision right" and ownership might simply mean "control." More concretely, the adaptation theory seems to apply not only to the make-or-buy problem but also (and perhaps better) to a particular class of contracting problems, where two firms with fixed boundaries pass decision rights across their boundaries by contract. These latter problems might be called "Contracting for Control," and a literature is emerging along these lines, as I discuss in the Conclusion.

2A. An Elemental Incentive-System Theory of the Firm

As described in Section I, the incentive-system theory can be seen as an accidental theory of the firm; the true focus of the incentive-system approach is on multi-task, multi-instrument agency problems. Accordingly, we begin by considering the following multi-task, *one*-instrument agency problem (*i.e.*, there is no asset yet, only an incentive contract).

Consider a technology of production $y = f_1a_1 + f_2a_2 + \varepsilon$ and a technology of performance measurement $p = g_1a_1 + g_2a_2 + \phi$, where a_1 and a_2 are actions chosen by the Agent and ε and ϕ are noise terms. Suppose that the parties are risk-neutral and the payoffs are $y - w$ to the Principal and $w - c(a_1, a_2)$ to the Agent, where w is the total compensation that the Principal pays the Agent, and suppose that the Agent's cost function is

$$(1) \qquad c(a_1, a_2) = \tfrac{1}{2}a_1^2 + \tfrac{1}{2}a_2^2.$$

If the Principal and Agent sign a linear contract $w = s + bp$, then the Agent's optimal actions are $a_1^*(b) = g_1b$ and $a_2^*(b) = g_2b$, and the resulting expected total surplus is $E(y) - c(a_1, a_2)$, or

$$(2) \qquad TS(b) = f_1 a_1^*(b) + f_2 a_2^*(b) - \tfrac{1}{2} a_1^*(b)^2 - \tfrac{1}{2} a_2^*(b)^2.$$

The efficient contract slope, b^*, maximizes this expected total surplus.[5]

---

[5] While b^* is the efficient slope for contracts of the form $w = s + bp$, one might wonder about contract with either different shapes or different arguments (or both). The linear contract shape can be

The central insight from this multi-task agency model is that the efficient slope depends on the alignment between the coefficient vectors (f_1, f_2) from y and (g_1, g_2) from p. At one extreme, if these vectors are perfectly aligned then the right choice of b can induce the Agent to choose the first-best values of a_1 and a_2; at the other extreme, if these vectors are orthogonal then the efficient value of b is zero (*i.e.*, the performance measure p is useless). See Feltham and Xie (1994), Datar, Kulp, and Lambert (2001), and Baker (2002) for richer models and further discussion.

To transform this agency problem into a theory of the firm, suppose there is a machine that the Agent uses in producing y. The resale value of the machine (after it is used in production) is $v = h_1a_1 + h_2a_2 + \xi$. The first-best actions by the Agent maximize the expected total surplus $E(y + v) - c(a_1, a_2)$, so we have $a_1^{FB} = f_1 + h_1$ and $a_2^{FB} = f_2 + h_2$. If the Principal owns the machine (so the Agent is an "Employee"), then the Principal's payoff is $y + v - w$ and the Agent's is $w - c$. The Agent's optimal actions are again $a_{1E}^*(b) = g_1b$ and $a_{2E}^*(b) = g_2b$, but now the efficient slope, b_E^*, maximizes a new expected total surplus, $E(y+v) - c(a_1, a_2)$, or

$$(3) \qquad TS_E(b) = (f_1 + h_1)\, a_{1E}^*(b) + (f_2 + h_2)\, a_{2E}^*(b) - \tfrac{1}{2}\, a_{1E}^*(b)^2 - \tfrac{1}{2}\, a_{2E}^*(b)^2.$$

Alternatively, if the Agent owns the machine (so the Agent is a "Contractor"), then the Principal's payoff is $y - w$ and the Agent's is $w + v - c$. Now the Agent's optimal actions are $a_{1C}^*(b) = g_1b + h_1$ and $a_{2C}^*(b) = g_2b + h_2$, and the efficient slope, b_C^*, maximizes the expected total surplus

$$(4) \qquad TS_C(b) = (f_1 + h_1)\, a_{1C}^*(b) + (f_2 + h_2)\, a_{2C}^*(b) - \tfrac{1}{2}\, a_{1C}^*(b)^2 - \tfrac{1}{2}\, a_{2C}^*(b)^2.$$

Simply put, having the Agent own the asset causes the Agent to respond to a given contract slope (b) differently than when the Agent does not own the asset, so the make-or-buy question becomes: which of the Agent's best-response functions – that of the employee, $(a_{1E}^*(b),\ a_{2E}^*(b))$, or that of the independent contractor, $(a_{1C}^*(b),\ a_{2C}^*(b))$ –

---

interpreted via the aggregation argument from Holmstrom and Milgrom (1987); see Gibbons (2004) for a simplified account. The dependence on p, rather than on y, is simply assumed: y is not contractible. Finally, one might wonder about selling y outright to the Agent, even if y is not contractible. Again, this possibility is simply assumed away – say, because the payoff y is not tied to any particular asset that could be sold to the Agent, but instead is a payoff that accrues to the Principal as an inextricable part of other (unmodeled) operations that the Principal owns.

allows the parties to achieve greater total surplus? As stark illustrations of optimal asset ownership in this incentive-system model, consider the following pair of examples:

Example 1: $y = a_1$, $v = a_2$, and $p = a_1 + a_2$.

Example 2: $y = a_1$, $v = a_2$, and $p = a_1$.

In Example 1, p is perfectly aligned with y + v, so the Principal can induce the Agent to choose first-best actions if the Principal owns the asset and the Agent (who is then an Employee) receives a contract with slope $b_E^* = 1$. In contrast, in Example 2, p is perfectly aligned with y, so the Principal can induce the Agent to choose first-best actions if the Agent owns the asset and the Agent (who is then a Contractor) receives a contract with slope $b_C^* = 1$. In Example 1, all the Agent's incentives come from the contract; in Example 2, part of the incentives come from the contract and the rest from asset ownership.

Richer versions of the incentive-system theory address additional instruments in the incentive problem. For example, job design can be modeled as changing the Agent's feasible action set: can the Agent work at home, or only at work; can the Agent make personal phonecalls from work; and so on. These enrichments are consistent with the basic theme of the incentive-system approach: asset ownership is useful when it gives the Principal improved control over the Agent's incentives.

2B. An Elemental Adaptation Theory of the Firm

As described in Section 1, the adaptation theory of the firm asks whether integration or non-integration better facilitates adaptive decision-making in environments where uncertainty is resolved over time. As a half-step towards such an adaptation theory of the firm, we begin with Simon's (1951) adaptation theory of employment, which we can now state formally. Two parties choose between (a) negotiating a decision ($d \in D$) before uncertainty ($s \in S$) is resolved, or (b) allocating authority to one party (the "boss"), who can then make a self-interested decision after s is observed. Payoffs are $U_B(s, d)$ to the boss and $U_E(s, d)$ to the employee, gross of any monetary transfers between the parties. Ex ante negotiation would therefore yield the decision d* that solves

$$(5) \qquad \max_{d \in D} E_s\{U_B(s,d) + U_E(s,d)\} ,$$

whereas giving the boss decision-making authority would yield the decision rule $d_B(s)$ that solves

$$(6) \qquad \max_{d \in D} U_B(s,d) .$$

Simon provides conditions on the payoff functions and the distribution of s such that giving authority to the boss is Pareto-superior to locking in an unconditional decision:

$$(7) \qquad E_s\{U_B(s, d_B(s)) + U_E(s, d_B(s))\} > E_s\{U_B(s, d^*) + U_E(s, d^*)\} .$$

Simon's paper is typically described as focusing on this analysis of when giving authority to the boss is Pareto-superior to locking in an unconditional decision. This issue was indeed Simon's main focus, but I think it is quite important that Simon also briefly discussed an elaboration of this model, in which the employee can be given the decision right instead of the boss (p. 304). That is, Simon envisioned the parties choosing among not only the unconditional decision d* but also the decision rules $d_B(s)$ and $d_E(s)$, where the latter solves

$$(8) \qquad \max_{d \in D} U_E(s,d) .$$

As we will now see, this elaboration of Simon's basic model is the key to an adaptation theory of the firm.

Although Simon's model laid valuable groundwork, it had one important drawback: if the parties can negotiate a decision ex ante, they presumably can renegotiate a decision ex post, so they need not settle for $d_B(s)$, but instead can renegotiate to the first-best decision $d^{FB}(s)$ that solves

$$(9) \qquad \max_{d \in D} U_1(s,d) + U_2(s,d) .$$

If the parties can always renegotiate to $d^{FB}(s)$, we no longer have either a theory of employment (comparing $d_B(s)$ to d*) or a theory of the firm (comparing $d_B(s)$ to $d_E(s)$).

Recent work – such as Aghion, Dewatripont, and Rey (2002) and Hart and Holmstrom (2002) – provides a way to resolve this lacuna in Simon's theory. This work analyzes an environment in which *decisions* are not contractible even ex post, but *decision rights* are contractible ex ante. As one way to motivate this distinction, imagine

that there is a moral-hazard problem ex post. More colorfully, let me recount a remark from an experienced academic / practitioner: "No contract in the world will cause an unwilling partner to perform."[6] Now, there may be some hyperbole in this remark, especially when applied to certain settings. On the other hand, I am convinced that in many settings this remark captures something important: this is why "working to rule" can be such a dramatic slowdown; more generally, this is why there can be such a gap between "consummate" and "perfunctory" performance (Williamson, 1975: 69).[7]

In this sub-section, I will follow Baker, Gibbons, and Murphy (2004a, 2004b) by assuming not only that decisions are not contractible ex post but also that *decision rights* cannot be renegotiated ex post (say, because the opportunity to take a decision after the state is realized is fleeting). The timing of the elemental adaptation theory of the firm is then as follows: (i) two parties negotiate over control of a decision right; (ii) the parties observe the state of the world, s, drawn from the finite set S according to the distribution p(s); (iii) the party with control chooses a decision, d, from the finite set D; (iv) the parties receive their payoffs, $U_i(s, d)$, for i = 1, 2. In short, we now have a model with one less option that Simon envisioned in his elaboration of the employment model: now the parties can choose between the decision rules $d_B(s)$ and $d_E(s)$, but they cannot negotiate a decision such as d* ex ante.

The analysis of this adaptation theory of the firm is extremely simple. If party i controls the decision right, then in state s party i will choose the decision $d^i(s)$ that maximizes $U_i(s, d)$, as in (6) and (8). The expected total surplus when party i controls the decision right is then

$$(10) \quad TS^i \equiv E_s\{U_1(s, d^i(s)) + U_2(s, d^i(s))\}.$$

This expected total surplus will typically not be first-best, as would arise if the parties could arrange to choose the decision rule $d^{FB}(s)$. Instead, in the negotiation in stage (i),

---

[6] Michael Levine, personal communication.

[7] A different way to resolve this lacuna in Simon's theory – one he might well have offered had the issue been raised – is to assume that opportunities to renegotiate a decision are fleeting. For example, imagine that the state is revealed at a random moment and that the decision must be taken just after the state is revealed or all value is lost. Such an assumption may well be plausible in some settings, but my guess is that the "moral hazard ex post" approach has both broader applicability and broader consequences for rethinking contract theory more generally.

the parties will allocate control of the decision right (perhaps in exchange for a side-payment) to the party who maximizes the expected total surplus in (10).

In sum, I see this elemental adaptation theory of the firm as a coherent theory without specific investments. The fact that such a theory was inspired, in significant part, by Williamson's 1975 book led me to argue in Section 1D that Williamson's collected works suggest two elemental theories of the firm – rent-seeking and adaptation. As further support for this claim, compare the indexes of his 1975 and 1985 books. The 1985 index lists many references to terms related to rent-seeking – 39 to "asset specificity," 7 to "dedicated assets," 5 to "firm-specific assets," 2 to "specialized assets," and 11 to "transaction-specific assets," and so on. In contrast, the 1975 index lists no references to any of these terms.

Of course, this difference in index references might reflect the evolution of terminology rather than the evolution of ideas. For example, the 1975 book does emphasize ideas that are *consistent with* specific investments and asset specificity, such as the transformation from large-numbers exchange ex ante to small-numbers exchange ex post. My claim, however, is that the 1975 book contains the germ of a theory that *can be independent of* certainly specific investments and perhaps asset specificity, as follows.

The elemental adaptation theory presented in this sub-section certainly does not involve specific investments (and, in the "contracting for control" interpretation sketched in the introduction to this section and expanded upon in the Conclusion, it need not involve assets, either). Furthermore, Williamson (1996: 377) defines asset specificity as a "specialized *investment* that cannot be redeployed ... except at loss of productive value" (emphasis added). On this definition, if the adaptation theory has no specific investments, then it would appear not to involve asset specificity. On the other hand, the adaptation theory does involve small-numbers bargaining and "bilateral dependency" (1996: 377); in this sense, the theory might be said to involve asset specificity, but it could just as well be said to involve "externalities." I am wary of using the term "asset specificity" in describing the adaptation theory if, in this context, that term is synonymous with time-honored terms such as externality. At a minimum, we can say that there is a coherent elemental theory of the firm without specific investments.

2C. An Elemental Property-Rights Theory of the Firm

As described in Section 1, the property-rights theory was the first formal theory to deliver a unified account of the costs and benefits of integration (*i.e.*, run the Coasean horserace on a level playing field). But in presenting three elemental formal theories in this section, I have departed from the historical order in which the theories were developed, for ease of exposition. In particular, the incentive-system and adaptation theories were simpler to explain than the property-rights theory is, because the property-rights theory incorporates aspects of both of the other theories, as follows.

Like the incentive-system theory, the property-rights theory allocates asset ownership to affect ex ante incentives. But the property-rights theory differs from the incentive-system theory by omitting the latter's non-contractible asset value (v) and contractible performance measure (p), so that the only ex ante incentives in the property-rights theory arise via control over asset utilization ex post. Like the adaptation theory, the property-rights theory envisions asset ownership as a pure decision right, with no direct effect on any party's utility function. But the property-rights theory differs from the adaptation theory by assuming that decisions are contractible ex post (and hence renegotiable), so that ex post asset utilization is (conditionally) efficient, regardless of asset ownership.

The timing of the elemental property-rights theory is thus as follows: (i) two parties negotiate over control of a decision right; (ii) the parties simultaneously choose actions, with party i choosing action $a_i \in A_i$ at cost $c_i(a_i)$; (iii) the parties observe both the actions (a_1, a_2) and the state of the world, $s \in S$; (iv) the parties negotiate over which decision, $d \in D$, the party with control should choose; (v) the parties receive their payoffs, $U_i(a_1, a_2, s, d)$ for $i = 1, 2$ (where U_i is gross of any monetary transfers and action costs). The actions (a_1, a_2) and the payoffs (U_1, U_2) are non-contractible. The decision d is non-contractible in stage (ii), but (unlike the adaptation model) becomes contractible in stage (iv), after the state is revealed in stage (iii). In sum, the basic argument in the property-rights theory is that the allocation of control in stage (i) determines the allocation of surplus in stage (iv), which determines investment incentives in stage (ii).

If the parties observe actions (a_1, a_2) and state s in stage (iii), then the (conditionally) efficient decision in stage (iv) solves

$$(11) \qquad \max_{d \in D} U_1(a_1, a_2, s, d) + U_2(a_1, a_2, s, d) \ ;$$

denote the solution by $d^*(a_1, a_2, s)$ and the resulting payoffs by $U_i^*(a_1,a_2,s) \equiv U_i(a_1, a_2, s, d^*(a_1, a_2, s))$. But if party i controls the decision right then, absent negotiation in stage (iv), party i will choose the decision $d^i(a_1, a_2, s)$ that solves

$$(12) \qquad \max_{d \in D} U_i(a_1,a_2,s,d) ,$$

resulting in the payoffs $U_j^i(a_1,a_2,s) \equiv U_j(a_1, a_2, s, d^i(a_1, a_2, s))$. Unless the parties have identical preferences over decisions, we will have $U_1^*(a_1,a_2,s) + U_2^*(a_1,a_2,s) > U_1^i(a_1,a_2,s) + U_2^i(a_1,a_2,s)$, so there will be an incentive for the parties to negotiate in stage (iv).

As a tractable model of negotiation when party i has control, suppose that the parties agree to the Nash Bargaining Solution, with the payoffs $(U_1^i(a_1,a_2,s),\ U_2^i(a_1,a_2,s))$ as the threat point. That is, the parties choose d and p to solve

$$(13) \qquad \max_{d \in D, p} \left[U_1(a_1,a_2,s,d) + p - U_1^i(a_1,a_2,s)\right]\left[U_2(a_1,a_2,s,d) - p - U_2^i(a_1,a_2,s)\right] ,$$

where p is the negotiated payment from party 2 to party 1 for which party i agrees to choose decision d.[8] The first-order condition for p yields

$$(14) \qquad p = \frac{1}{2}\left[\left(U_2(a_1,a_2,s,d) - U_2^i(a_1,a_2,s)\right) - \left(U_1(a_1,a_2,s,d) - U_1^i(a_1,a_2,s)\right)\right] ,$$

from which it follows that d solves

$$(15) \qquad \max_{d \in D} \frac{1}{4}\left[\left(U_1(a_1,a_2,s,d) - U_1^i(a_1,a_2,s)\right) + \left(U_2(a_1,a_2,s,d) - U_2^i(a_1,a_2,s)\right)\right]^2 ,$$

so $d = d^*(a_1, a_2, s)$. That is, the parties bargain to the efficient decision, conditional on the observed actions and state, regardless of asset ownership.

But ownership does matter, even if it does not affect decision-making in stage (iv), because the payoffs in stage (iv) depend on ownership. When party i has control, party j's net payoff (ignoring payments negotiated in stage (i)) is

---

[8] Note that p here is a payment decided through negotiation, rather than the performance measure used in the incentive contract in the incentive-system theory. I trust that the variable definitions will be clear from context, but I also note that these two definitions of p play similar roles (in shaping ex ante incentives) in the incentive-system and property-rights theories, so there is a sense in which I here encourage a small degree of notational confusion!

$$(16) \quad NP_j^i(a_1,a_2,s) \equiv \frac{1}{2}\left(U_j^*(a_1,a_2,s) + U_k^*(a_1,a_2,s)\right) + \frac{1}{2}\left(U_j^i(a_1,a_2,s) - U_k^i(a_1,a_2,s)\right) - c_j(a_j),$$

where i, j, k $\in \{1, 2\}$ and j $\neq$ k. Note that the first term involves the efficient total surplus, $U_j^*(a_1,a_2,s) + U_k^*(a_1,a_2,s) \equiv ETS(a_1,a_2,s)$, whereas the second involves the threat-point differential, $U_j^i(a_1,a_2,s) - U_k^i(a_1,a_2,s) \equiv TPD_j(a_1,a_2,s)$. In terms of this new notation, (16) says that, in choosing an action in stage (ii), party j has a half-strength incentive to maximize ETS(a_1, a_2, s), but also a half-strength incentive to maximize TPD$_j$(a_1, a_2, s). In particular, from the threat-points term, party j has incentives not only to improve her threat point but also to worsen party k's.[9]

We can now analyze whether expected total surplus is higher if party 1 owns the asset or party 2. From this governance-choice perspective, the half-strength incentives to maximize the efficient total surplus are irrelevant: these incentives exist regardless of who owns the asset, so *optimal asset ownership is determined entirely by the threat-points term*. In particular, we would like to find a governance structure such that the existing half-strength incentives from TPD$_j$(a_1, a_2, s) closely approximate the missing half-strength incentives to maximize ETS(a_1, a_2, s). How well we can succeed in this quest turns out to depend on the details of the model: the threat-points term could create incentives that are too week, too strong, or just right; furthermore, in a multi-task setting, these incentives could be well aligned with the maximized total surplus or misdirected.

Formally, given the negotiated decision (d) and payment (p) in stage (iv), we can solve for the actions induced in stage (ii), and hence for the optimal ownership structure in stage (i). If party i owns the asset then party j will choose the action a_j that solves

$$(17) \quad \max_{a_j \in A_j} \; E_s\left[NP_j^i(a_j, a_k^*, s)\right],$$

producing the Nash equilibrium actions (a_1^*, a_2^*), so expected total surplus will be

$$(18) \quad TS^i \equiv E_s\left(U_1^*(a_1^*, a_2^*, s) + U_2^*(a_1^*, a_2^*, s)\right) - c_1(a_1^*) - c_2(a_2^*) \; .$$

---

[9] Grout (1984) provided an early analysis along these lines, in the context of union bargaining, but for a fixed governance structure. That is, Grout showed how ex post bargaining creates ex ante incentives to manipulate the threat-points term, but he did not explore how different governance structures thus create different ex ante incentives.

Efficient negotiation in stage (i) will result in the governance structure that maximizes expected total surplus. In this simple setting, with only one asset, determining efficient asset ownership is simply a comparison of TS^1 to TS^2.

Whinston (2003) solves a linear-quadratic version of this model, which illustrates a point also raised by Holmstrom and Roberts (1998) and Holmstrom (1999): only the margins matter. That is, since our goal is to find a governance structure such that the existing half-strength incentives from $TPD_j(a_1, a_2, s)$ closely approximate the missing half-strength incentives to maximize $ETS(a_1, a_2, s)$, we can focus on the determinants of these threat-points incentives, which are (of course) the marginal effects of the parties' actions on $TPD_j(a_1, a_2, s)$; any constant term in $TPD_j(a_1, a_2, s)$ is irrelevant, as is any term that depends on s but not on either a_1 or a_2. Whinston then notes that margins are hard to observe (especially when they are the marginal effects of non-contractible actions on non-contractible payoffs), so direct tests of the property-rights theory may be difficult. Holmstrom and Roberts take a complementary approach, noting that a striking prediction of the property-rights theory is that the intercept and s-term of $TPD_j(a_1, a_2, s)$ should *not* matter for governance choice.

2D. Interim Summary

These elemental versions of the incentive-system, adaptation, and property-rights theories of the firm suggest a two-by-two representation of alternative models of ownership. The incentive-system theory (IS) illustrates that changing ownership can change the allocation of payoff rights, whereas the adaptation theory (Ad) illustrates that changing ownership can change the control of decision rights. Finally, the property-rights theory (PR) can be seen as a blend of the two, where a change in the control of decision rights induces a change in the net payoff function (16), which then functions very much like a change in the allocation of payoff rights in determining the parties' optimal ex ante actions. These distinctions are summarized in Table 1 below.

		PAYOFF RIGHTS	
		Changed	Fixed
DECISION RIGHTS	Changed	PR	Ad
	Fixed	IS	

Table 1: Payoff and Decision Rights in Three Theories of the Firm

Having thus compared formal versions of three recent theories of the firm, we can now return to the granddaddy of such theories – the rent-seeking theory of the firm – and ask how it compares to the other three.

3. An Elemental Rent-Seeking Theory of the Firm?

The rent-seeking theory of the firm seems never to have been fully formalized, at least in the sense of providing a unified account of the costs and benefits of integration. On the other hand, the theory certainly caught the attention of empirical researchers, who have now compiled a substantial body of evidence relating asset specificity to vertical integration (Boerner and Macher, 2002). I will therefore try to describe both what I believe the rent-seeking theory currently is and what I believe it could and should be. To do so, I proceed via four steps. First, I follow Masten (1982, 1986) in borrowing the basic model of rent-seeking from Tullock (1980) and reinterpreting it as a model of "post-contractual opportunism" between firms. Second, I discuss the need for costs of integration, in order to transform this basic model of rent-seeking between firms into a theory of the firm. Third, I follow a long tradition, dating at least to Knight (1921), by observing that rent-seeking also occurs within firms, not just between. Finally, I provide an elemental theory in which rent-seeking occurs both between and within firms.

3A. Rent-Seeking Between Firms

I will define rent-seeking as individually optimal (but socially destructive) haggling over appropriable quasi-rents. This definition is intended to be synonymous with Williamson's (1971: 114-5) use of "haggling" as quoted in Section 1: "fiat is frequently a more efficient way to settle minor conflicts ... than is haggling;" in some circumstances, non-integration cannot avoid inefficient haggling because, while "jointly (and socially) unproductive, it constitutes a source of private pecuniary gain," so integration (with dispute-resolution by fiat) can then be more efficient. This definition of rent-seeking is also intended to be synonymous with Klein-Crawford-Alchian's "post-contractual opportunism" and broadly consistent with Williamson's "opportunism" (self-interest-seeking with guile).

With all these terms already in the literature, one might question the wisdom of importing a new one. I use "rent-seeking" exactly because it has not only an established meaning in a collection of applications (Buchanan, Tollison, and Tullock, 1980) but also – and more importantly, for my purposes – an established model, due to Tullock (1980). In this basic model, two contestants (A and B) purchase lottery tickets in amounts x_A and x_B, respectively, each hoping to win a prize of \$100. Contestant i's expected payoff is then $x_i/(x_i + x_j) * 100 - x_i$, so the symmetric Nash equilibrium has positive expenditures on lottery tickets, but in this equilibrium the contestants have the same probabilities of winning as if neither had spent anything (namely, one-half). The subsequent literature (*e.g.*, Perez-Castrillo and Verdier, 1992; Rajan and Zingales, 2000; Skaperdas, 1992; Szidarovszky and Okuguchi, 1997) has generalized and applied this basic model.

Most of the applications of the rent-seeking model have not focused on the theory of the firm, but Masten (1982, 1986) used such a model to analyze the transaction costs of non-integration (*i.e.*, "haggling" or "post-contractual opportunism" between firms). In terms of the notation used in Section 2, Masten's model of non-integration can be expressed as

$$(19) \qquad U_i(d_i,d_j) = P_i(d_i,d_j) * AQR - c_i(d_i) \ ,$$

where U_i is the expected payoff to party i, P_i is the probability that party i wins the contest (where the prize is now an appropriable quasi-rent with value AQR), and c_i is the cost incurred by party i from taking decisions d_i in the attempt to win the contest. As in the adaptation theory in Section 2B, the decisions d are not contractible even ex post, so Nash equilibrium is the natural solution concept for analyzing (19). For example, given symmetry assumptions on the probability functions P_i and the cost functions c_i, the Nash-

equilibrium decisions are $d_i = d_j = d^{NE}$, so each party has probability one-half of capturing AQR, but each incurs costs $c(d^{NE}) > 0$. These rent-seeking costs, $2c(d^{NE})$, are inefficient. Furthermore, given standard regularity conditions, the size of this inefficiency increases with AQR.

This comparative-static result – that inefficiency increases with AQR – seems to have motivated much of the large empirical literature that relates asset specificity to vertical integration. Of course, (19) per se is not yet a theory of the firm, because so far (19) does not even define integration, not to mention show why integration can produce a tradeoff. Masten applies the natural definition of integration in the context of (19): since non-integration means that each party controls its own decision right, integration means that one party controls both decision rights and so can capture all of AQR at trivial cost. For example, suppose $P_i(d_i, d_j) = d_i / (d_i + d_j)$, as in Tullock's basic model. If party i controls both decision rights, then i can choose $d_i = \varepsilon$ and $d_j = 0$, thereby capturing the AQR ($P_i = 1$) at trivial cost ($c(\varepsilon) \sim 0$).

3B. Costs of Integration (in a Unified Theory of the Firm)

Having defined integration, we now need to explain why integration can produce a tradeoff. So far, the costs of non-integration are $2c(d^{NE}) > 0$, whereas the costs of integration are $c(\varepsilon) \sim 0$, so integration is always more efficient than non-integration. To produce a tradeoff, we need costs of integration. Williamson has discussed such costs for over 30 years (1971, p. 113; 1975, Ch. 7; 1985, Ch. 6; 1996, Ch. 4; 2002, p. 177) – often with persuasive descriptions of the potential ills of internal organization, sometimes summarized by labels such as "costs of bureaucracy." A blunt way to introduce such costs into the rent-seeking model is simply to assume that integration entails costs of bureaucracy, denoted by B, so that we must now compare $c(\varepsilon) + B$ to $2c(d^{NE})$ to determine whether integration or non-integration is more efficient. The question then becomes: what determines B?

One approach is to assume that the determinants of the costs of bureaucracy are independent of the determinants of the costs of rent-seeking. This assumption has the advantage that it motivates the logistic regressions that are common in the empirical literature that relates asset specificity to vertical integration, but this assumption also has theoretical, econometric, and empirical disadvantages that seem quite severe (at least to me). I describe the theoretical and econometric disadvantages in this sub-section and the empirical disadvantages in the next.

The theoretical disadvantage of assuming that the costs of bureaucracy are independent of the costs of rent-seeking is that such an assumption is inconsistent with developing a unified theory of the firm. I sometimes find this point tricky to convey, so let me begin this discussion where it will end: the issue is with the word "assuming," not with the word "independent." That is, if we are to have a unified theory of the firm in which the costs of bureaucracy are independent of the costs of rent-seeking, then we need a model that proves this independence, not an assumption that simply asserts it.

Figure 1 provides a graphical summary of a (simple) unified theory of the firm, in which the costs of integration are definitely not independent of the costs of rent-seeking. The figure illustrates not only Coase's (1937) famous hypothesis (that firms exist only where they perform better than markets would, which is to the right of the dotted vertical line in this figure), but also its unremarked corollary (that the firms we then observe will be less efficient than the markets we observe, even though the firms we observe will be more efficient than the markets they replaced). Because the unremarked corollary involves a classic sample-selection argument, I have elsewhere titled this figure "Coase (1937) Meets Heckman (1976)" (Gibbons, 2003: 757).

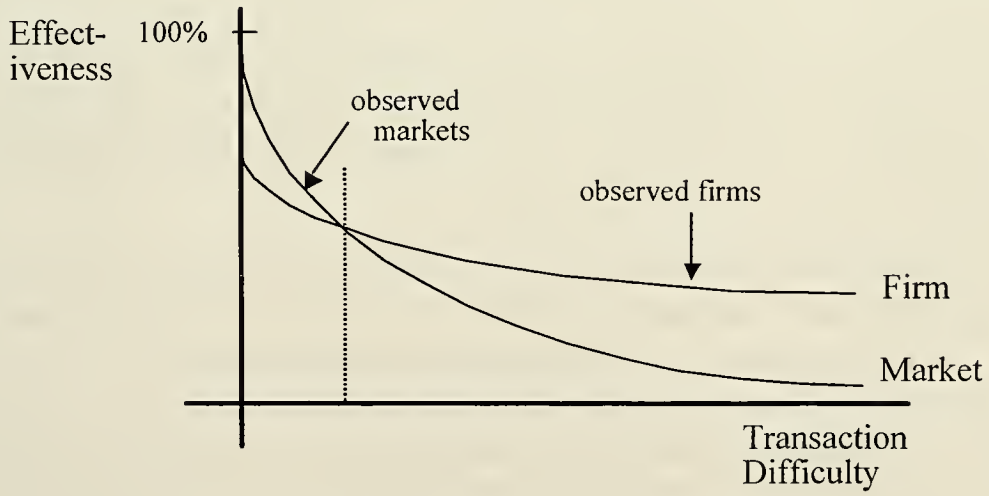

Figure 1: Coase (1937) Meets Heckman (1976)

Figure 1 implicitly asserts that forces sufficient to wreck market exchange cannot be assumed to disappear once transactions are internalized. This is the central point of any unified theory of the firm: any source of transaction costs under non-integration

could continue to cause problems under integration, and vice versa. More generally than in the simple theory depicted in Figure 1, which has only one source of "transaction difficulty," a unified theory of the firm derives both the costs and the benefits of integration from a fixed *set* of sources. Given a fixed set of sources of costs and benefits of integration, it is entirely possible that some sources in this set have greater influence on the costs of integration, while other sources have greater influence on the benefits. But it is unacceptable in a unified theory of the firm to allow any particular source to have an influence on (say) the costs of integration, while simply assuming that this source has no influence on the benefits.

Figure 1 is similar to figures such as Williamson's (1985) Figure 4-1, but three differences are worth discussing. To discuss these differences, I first describe Williamson's figure. In developing Figure 4-1, Williamson follows Riordan and Williamson (1985) by defining k to be an index of asset specificity, $\beta(k)$ to be the bureaucratic costs of internal governance, $M(k)$ to be the corresponding governance costs of markets, and $\Delta G(k) = \beta(k) - M(k)$ to be the cost advantage of markets over firms for transactions with asset specificity k. He then argues that $\beta(0) > M(0)$ (*i.e.*, markets have lower governance costs at zero asset specificity), but that $\beta'(k) < M'(k)$ at each value of k (*i.e.*, market governance costs increase faster, at every level of asset specificity). If one adds a mild regularity condition, it then follows that, plotting $\Delta G(k)$ as a function of k, there is a critical value k^* that is analogous to the dotted vertical line in Figure 1, in the sense that $\Delta G(k) > 0$ if and only if $k < k^*$ (*i.e.*, markets are the efficient governance structure for $k < k^*$, but firms are efficient for $k > k^*$).

The first difference between Figure 1 and Williamson's figure is that Figure 1 allows alternative sources of transaction difficulty, beyond asset specificity. For example, Figure 1 can be interpreted in terms of the incentive-system theory described in Section 2A (which makes no mention of asset specificity), as follows: set $\underline{f} = (1, 0)$, $\underline{h} = (0, 1)$, and $\underline{g} = (\cos\theta, \sin\theta)$, and interpret $\theta \in [0, \pi/4]$ as the measure of transaction difficulty. For this parameterization of the incentive-system model, $TS_E > TS_C$ if θ is sufficiently large; specifically, the dotted vertical line occurs where $2\sin\theta = \cos\theta$. The fact that Figure 1 allows alternative sources of transaction difficulty is consistent with this essay's focus on four theories of the firm, not just on the rent-seeking theory (and, in particular, not just on asset specificity).

The second difference is that Figure 1 can be derived from formal micro-foundations, as just given for the incentive-system theory. The advantage of such a

micro-foundations approach to interpreting Figure 1, relative to the reduced-form approach that Williamson takes in developing Figure 4-1, is not that this approach is quantitative (*e.g.*, it produces a specific value of θ at which the dotted vertical line occurs), but rather that this approach is complete (*i.e.*, it must *derive* cost functions akin to $M(k)$ and $\beta(k)$, rather than imposing assumptions about them). In particular, in Williamson's context of asset specificity, arguments like those given in Section 3A could be used to derive that $M'(k) > 0$, but that discussion of rent-seeking between firms is no help (that I can see) in deriving $\beta'(k)$. The fact that Figure 1 can be derived from micro-foundations is consistent with this essay's focus on formal theories.

Finally, the third difference is that Figure 1 generalizes Figure 4-1 by plotting cost functions akin to $\beta(k)$ and $M(k)$ separately and asking where they cross, rather than plotting the cost difference $\Delta G(k)$ and asking where it crosses the k-axis. This may seem to be a trivial point, because both approaches produce the same value of k^*, but plotting the cost difference $\Delta G(k)$ hides the cost levels that are necessary to compare the efficiency of observed markets to that of observed firms. That is, Williamson's figure captures the Coase part of Figure 1 but not the Heckman part, and the latter is important in the discussion of the econometric disadvantages of the independence assumption, to which we now turn.[10]

For econometric purposes, it might seem convenient (and relatively inconsequential, at least as a first approximation) to assume that the costs of bureaucracy are independent of the costs of rent-seeking (*i.e.*, $\beta'(k) = 0$). Many papers explicitly or implicitly impose this assumption and then estimate organizational form as a function of asset specificity. But Masten, Meehan, and Snyder (1991, hereafter MMS) note that if asset specificity affects the transaction costs of *both* integration and non-integration then "a finding that higher asset specificity leads to a larger probability of integration could, in principle, obtain even if the hypothesis that asset specificity raises [the transaction costs of non-integration] were invalid" (p. 4). That is, one could find that $\Delta G'(k) < 0$ even if the hypothesis that $M'(k) > 0$ were invalid, provided that $\beta'(k) < 0$. As a result, the standard approach – which amounts to testing whether $\Delta G'(k) < 0$ – is "unable to distinguish whether observed patterns of organization resulted from hypothesized

[10] Figure 1 in Williamson (1991) does plot the cost functions separately, rather than plotting their difference, and so could be used to analyze the econometric issues discussed below, but Williamson does not use this figure for that purpose.

changes in market transaction costs or from systematic, but as yet unexplored, variations in the costs incurred organizing production internally" (p. 2).

Of course, MMS's econometric concerns apply not just to asset specificity but also to any transaction characteristic that could affect the performance of both integration and non-integration – that is, to any theory of the kind depicted in Figure 1. Thus, having been motivated by the hypothesis that the transaction costs of non-integration increase with factors such as asset specificity, and having established a correlation between vertical integration and such factors, it is time for the empirical literature to recognize that firms are unlikely to be immune to the forces that wreck markets.

3C. Rent-Seeking Within Organizations

I am fully persuaded that rent-seeking between organizations is an important transaction cost of non-integration. But after a detailed and persuasive discussion of such "post-contractual opportunism" under non-integration, Klein, Crawford, and Alchian (KCA) observe that: "Vertical integration does not completely avoid contracting problems. The firm could usefully be thought of as a complex nonmarket contractual network where *very similar forces* are present" (p. 299, emphasis added). KCA go on to quote Frank Knight as stating: "[T]he internal problems of the corporation, the protection of its various types of members and adherents against each other's predatory propensities, are quite as vital as the external problem of safeguarding the public interests against exploitation by the corporation as a unit." In the same vein, but outside economics, the large literature on power and politics in organizations – from March (1962) and Cyert and March (1963) through Pfeffer (1981) and well beyond – has devoted substantial attention to a host of issues related to those raised by Knight and KCA. In short, politicking within firms seems to be the inescapable internal-organization analog of haggling between firms.

If rent-seeking does not disappear once transactions are internalized, there is also an empirical disadvantage (not just theoretical and econometric disadvantages) to a theory of the firm that emphasizes the costs of rent-seeking between firms but closes the model by assuming the existence of independent costs of bureaucracy. It seems preferable to develop an elemental rent-seeking theory of the firm, where rent-seeking provides a unified account of both the costs and the benefits of integration.

One way to develop such an elemental rent-seeking theory of the firm is to enrich equation (19), as follows. Suppose that, in addition to the decision rights d_1 and d_2 that can be controlled by either party, there are also decision rights λ_i and λ_j that can be controlled only by parties i and j, respectively. The alienable decision rights d_1 and d_2 might be the rights to control two pieces of physical capital, whereas the inalienable decision rights λ_i and λ_j might be the rights to control the parties' own human capital. As in (19), non-integration means that each party controls one of the alienable decision rights, whereas integration means that one party controls both alienable decision rights, but now each party controls its own inalienable decision right under both integration and non-integration. Finally, as was true of d in (19), now neither d nor λ is contractible even ex post.

For the non-integrated case where party i controls d_i, we could rewrite (19) as

$$(20) \qquad U_i(d_i,d_j,\lambda_i,\lambda_j) = P_i(d_i,d_j,\lambda_i,\lambda_j) * AQR - c_i(d_i) - k_i(\lambda_i) \ .$$

Under symmetry assumptions on P_i, c_i, and k_i, non-integration yields $d_i = d_j = d^{NE}$ and $\lambda_i = \lambda_j = \lambda^{NE}$, so each party again has probability one-half of capturing AQR but incurs costs $c(d^{NE}) + k(\lambda^{NE}) > 0$. The difference between (19) and (20) is that now integration does not mean that one party controls all the relevant decision rights. Instead, even under integration (of the alienable decision rights), both parties use whatever decision rights they control (including at least their inalienable decision rights) to attempt to capture AQR. There is thus rent-seeking not only between organizations but also within.

Of course, (20) is so abstract as to be almost useless. To refine this approach, one must give plausible descriptions of the alienable and inalienable decision rights, (d_i, d_j) and (λ_i, λ_j). Furthermore, one must show when and why integration produces a tradeoff. Milgrom and Roberts (1988, 1990) have sketched a theory of the firm along roughly these lines. In their first paper, they offer a rich description and model of not only how "influence activities" occur in organizations but also why organizations choose designs that do not eliminate such rent-seeking. And in their second paper, Milgrom and Roberts argue that influence activities are a leading cost of integration (*i.e.*, giving someone authority means that she will get lobbied) and then sketch a theory of the firm in which the costs of non-integration follow from "bargaining costs" between firms.

The Milgrom-Roberts approach seems promising. In particular, taking influence activities to be an example of rent-seeking within organizations, this approach could

produce an elemental rent-seeking theory of the firm, if "bargaining costs" can be formalized in a way that captures the idea of rent-seeking between firms. In addition, because rent-seeking / influence activities seem to exist both within and between firms, this approach could produce a unified theory of the firm, where a fixed set of transaction difficulties produce both the costs and the benefits of integration. I attempt to develop such a model in the next sub-section.[11]

3D. An Elemental Influence-Cost Theory of the Firm?

I now develop a simple model of influence activities within and between organizations, using a model related to Holmstrom's (1982) model of career concerns, Fudenberg and Tirole's (1986) model of signal jamming, and Milgrom and Roberts's (1988) model of influence activities. The first goal of the model is to derive the following intuitive comparative-static result: influence costs are higher when the influencer cares more about the decision. The second goal of the model is to see whether this comparative-static result could be the driving force in a theory of the firm, akin to the result from (19) that the inefficiency of non-integration increases with AQR.

The timing of the model is as follows: (i) two parties negotiate over control of a single alienable decision right; (ii) the parties simultaneously choose "lobbying" activities, $\lambda_i \in \Lambda$ at cost $k(\lambda_i)$; (iii) the parties publicly observe the signal σ; (iv) the party with control chooses a decision, $d \in D$; and (v) the parties receive their payoffs, $U_i(s, d)$ for $i = 1, 2$ (where U_i is gross of any monetary transfers and lobbying costs). This model blends aspects of the adaptation model from Section 2B with equation (20) from Section 3C. In particular, this model enriches the adaptation model by adding the lobbying activities in stage (ii), which can be interpreted as the inalienable decision rights (not coincidentally also denoted by λ_i and λ_j) in (20).

Let the payoff functions be $U_i(s, d) = - \alpha_i (d - s - \beta_i)^2$, where $\alpha_i > 0$ and $\beta_i \in \Re$. The parameter α measures the party's sensitivity to the difference between the decision taken and that party's ideal decision; the parameter β measures how the party's ideal decision differs from the state. The desired comparative-static result – that influence costs are higher when the influencer cares more about the decision – can then be stated as follows: when party i controls the decision right, the equilibrium value of $k(\lambda_j)$ increases with α_j.

[11] Meyer, Milgrom, and Roberts (1992) also develop a model of roughly this kind, but focused more on influence costs within firms (and the implications for divestitures and corporate focus) than on haggling between firms.

Let the public signal observed in stage (iv) be $\sigma = s + \lambda_i + \lambda_j + \varepsilon$, and let there be symmetric uncertainty: both parties share the prior belief that s is Normally distributed with mean m and variance 1/h (*i.e.*, precision h) and ε is Normally distributed with mean zero and precision h_ε. Because the state s is Normal, let $D = \Re$. Similarly, let $\Lambda = \Re$, so that each party can move σ either up or down. Finally, let the cost function be symmetric around zero, with $k'(0) = 0$, $k'(-\infty) = -\infty$, $k'(\infty) = \infty$, and $k'' > 0$.

The public signal σ is the crux of the career-concerns / signal-jamming / influence-activities model: even though neither party knows the true state, both parties care about how the eventual decision will relate to the state, so the party with control will try to extract from the signal whatever information σ might contain about the state, so the other party will try to move the realization of σ in a direction that is favorable to her. In equilibrium, both parties have correct beliefs. In particular, in equilibrium, the party with control correctly anticipates the other party's attempts to influence σ, and so correctly accounts for those attempts in interpreting σ as a signal about s. As we will see, however, the other party still has an incentive to influence σ. To put this (somewhat slippery) point differently, the equilibrium level of lobbying cannot be zero, because if the party with control believed the other party to be doing no lobbying, then the other party would have a strong incentive to lobby.

To analyze the model, suppose that party i has control. Then in stage (iv) she will choose d to solve

$$(21) \qquad \max_{d \in D} E_s[U_i(s,d) \mid \sigma] ,$$

so the solution is $E_s[s \mid \sigma] + \beta_i$. If party i's conjecture about j's lobbying is $\hat{\lambda}_j$, then

$$(22) \qquad E_s[s \mid \sigma] = \frac{hm + h_\varepsilon(\sigma - \lambda_i - \hat{\lambda}_j)}{h + h_\varepsilon} .$$

Equation (22) is one of two key aspects of a career-concerns / influence-activities / signal-jamming model, because it shows how the party with control attempts to account for the other party's lobbying activities, via the conjecture $\hat{\lambda}_j$.

Given i's posterior belief about s after observing σ, i's decision will be

$$(23)\quad d_i^*(\sigma,\hat{\lambda}_j) \equiv \frac{hm + h_\varepsilon(\sigma - \hat{\lambda}_j)}{h + h_\varepsilon} + \beta_i ,$$

where $\lambda_i = 0$ because $k(0)$ minimizes $k(\lambda)$. Working backwards to stage (ii), party j will therefore chose λ_j to solve

$$(24)\quad \max_{\lambda_j \in \Lambda} -\alpha_j E_{s,\varepsilon}[(d_i^*(\sigma,\hat{\lambda}_j) - s - \beta_j)^2] - k(\lambda_j) .$$

Because

$$(25)\quad d_i^*(\sigma,\hat{\lambda}_j) - s - \beta_j = \frac{h_\varepsilon}{h + h_\varepsilon}\lambda_j + \left\{ \frac{h}{h + h_\varepsilon}(m - s) + \frac{h_\varepsilon}{h + h_\varepsilon}(\varepsilon - \hat{\lambda}_j) - B_j \right\} ,$$

where $B_j \equiv \beta_j - \beta_i$, we have

$$(26)\quad E_{s,\varepsilon}[(d_i^*(\sigma,\hat{\lambda}_j) - s - \beta_j)^2] = \left(\frac{h_\varepsilon}{h + h_\varepsilon}\right)^2 \lambda_j^2 - 2\lambda_j \left(\frac{h_\varepsilon}{h + h_\varepsilon}\right)\left(\frac{h_\varepsilon}{h + h_\varepsilon}\hat{\lambda}_j - B_j\right) + M ,$$

where M collects the terms not involving λ_j. The first-order condition for λ_j is therefore

$$(27)\quad -\alpha_j \left\{ 2\left(\frac{h_\varepsilon}{h + h_\varepsilon}\right)^2 (\lambda_j - \hat{\lambda}_j) - \frac{h_\varepsilon}{h + h_\varepsilon} B_j \right\} = k'(\lambda_j) ,$$

which implicitly defines $\lambda_j^*(\hat{\lambda}_j)$, j's best response to i's conjecture. Equation (27) is the second key aspect of a career-concerns / influence-activities / signal-jamming model, because it shows how the party without control optimally responds to the conjecture $\hat{\lambda}_j$ held by the party with control.

In equilibrium, i's conjecture must be correct, so imposing $\lambda_j^*(\hat{\lambda}_j) = \hat{\lambda}_j$ in (27) yields the first-order condition for the equilibrium level of j's lobbying activity. Denoting this equilibrium level of lobbying by λ_j^*, we have

$$(28)\quad 2\frac{h_\varepsilon}{h + h_\varepsilon}\alpha_j B_j = k'(\lambda_j^*) .$$

We therefore have the comparative-static result that was the first goal of this model: $k(\lambda_j^*)$ increases with α_j.

To explore the implications of this comparative-static result for a theory of the firm, we work backwards to stage (i). If party i controls d then influence costs will be $k(\lambda_j^*)$. Alternatively, if party j controls d then influence costs will be $k(\lambda_i^*)$. Because $B_j = -B_i$, we have that $k(\lambda_j^*) > k(\lambda_i^*)$ if and only if $\alpha_j > \alpha_i$. But this comparison of influence costs is not the only determinant of the efficient allocation of control; we must also consider the decision-making by the party in control, as follows.

Allocating control to party i will produce an expected total payoff of

$$(29) \quad -\alpha_i E_{s,\varepsilon}[(d_i^*(\sigma,\lambda_j^*) - s - \beta_i)^2] - \alpha_j E_{s,\varepsilon}[(d_i^*(\sigma,\lambda_j^*) - s - \beta_j)^2] - k(\lambda_j^*) ,$$

where, in equilibrium,

$$(30) \quad d_i^*(\sigma,\lambda_j^*) \equiv \frac{hm + h_\varepsilon(s+\varepsilon)}{h + h_\varepsilon} + \beta_i .$$

The expected total payoff from allocating control to party i is therefore

$$(31) \quad L_i = -\alpha_j B_j^2 - k(\lambda_j^*) - L_0 ,$$

where L_0 collects the terms that are independent of the allocation of control. Thus, the efficient allocation of control is to the party with the higher value of α, but this result holds for two reasons, only one of which was the second goal of this model. That is, giving control to the party with the higher value of α minimizes not only the second term in (31), which reflects the loss from lobbying activities, based on the comparative-static result that was the first goal of this model, but also the first term, which reflects the loss from self-interested decision-making by party i, just as would occur in an adaptation model without lobbying activities.

If an elemental influence-cost theory of the firm is distinct from an elemental adaptation theory of the firm, then the two should not produce identical prescriptions concerning the efficient allocation of control. It may be that richer models in the spirit of this simple influence-cost model – such as with two alienable decision rights, d_1 and d_2, as in (19) and (20) – can avoid this problem of identical prescriptions, but this essay is not the place to pursue richer models. I think this approach is promising because it could deliver a unified rent-seeking / influence-cost theory of the firm. Such a theory would not only recognize that haggling / politics are inescapable problems in any governance structure, it would also explain why integration manages these problems better than non-integration does for some parameters but not for others.

3E. Summary

I have tried to describe (somewhat implicitly) both what I believe the rent-seeking theory currently is and what I believe it could and should be. Since parts of this argument were rather involved, let me recapitulate it here.

Tullock's (1980) basic model of rent-seeking has the attractive feature that inefficiency increases with the size of the prize. This model offered an appealing way for Masten (1982, 1986) to formalize Williamson's (1979) argument that the transaction costs of non-integration increase with asset specificity. But Tullock's model formalizes only a theory of the transaction costs of non-integration, not a theory of the firm, because the model does not produce a tradeoff between non-integration and integration.

To produce a tradeoff, we need costs of integration. It is common (especially in empirical work) to invoke costs of internal organization, often labeled something like "costs of bureaucracy." If one assumes that these costs of internal organization are independent of the rent-seeking costs of non-integration, then one motivates the logistic regression that is familiar in the literature that relates asset specificity to vertical integration. But this independence assumption has both theoretical and econometric disadvantages. Theoretically, simply asserting that the determinants of the costs of bureaucracy have no impact on the costs of rent-seeking (and vice versa) is inconsistent with developing a unified theory of the firm. Econometrically, running the familiar logistic regression cannot discriminate between two very different hypotheses – that asset specificity (or any other source of transaction difficulty) increases the transaction costs of non-integration, versus that asset specificity decreases the costs of internal organization. The fact that we cannot yet generate arguments in support of the latter hypothesis does not mean that we should employ econometric approaches that prevent the data from telling us that the latter hypothesis might be correct.

In addition to these theoretical and econometric disadvantages of assuming that the costs of bureaucracy are independent of the costs of rent-seeking, there is also an empirical disadvantage: there is abundant theoretical and empirical pressure – from Knight (1921), Klein-Crawford-Alchian, and others inside economics, and from March (1962), Pfeffer (1981) and others outside – to suspect that rent-seeking occurs within firms as well as between. It would therefore be very appealing to have a rent-seeking theory of the firm that fits the "Coase Meets Heckman" argument in Figure 1. In such a theory, firms would not be immune to forces that destroy markets. Instead, as the environment becomes more susceptible to rent-seeking (in some sense to be defined), the

performance of both integration and non-integration would decline, but the theory would explain when and why integration is superior to non-integration. Such a theory would not only have theoretical appeal; it would also shed light on how to construct econometric tests of the effect of rent-seeking on organizational form.

Let me conclude this section's discussion of the rent-seeking theory of the firm with two simple observations: first, rent-seeking within organizations appears to be an empirical fact; second, a unified rent-seeking theory of the firm is therefore an attractive theoretical possibility. In holding out this goal of a unified theory of the firm, I am not espousing an aesthetic position held by a tiny band of abstract theorists. To the contrary, I am reiterating one of Williamson's earliest visions for organizational economics: the more we come to understand about the forces that wreck market exchange, Williamson argued, the more we require "a parallel treatment of the sources and consequences of failures of internal organization as they relate to vertical integration" (1971: 122); furthermore, he later argued, "I submit ... that substantially the same factors that are ultimately responsible for market failures also explain failures of internal organization" (1973: 316).

4. An Integrative Framework (for Differentiating Among Theories)

In this section I develop a framework that nests the four elemental theories. I then locate each theory within this integrative framework. Locating the four theories shows that the property-rights and incentive-system theories are closely related, but that these theories are quite different from the rent-seeking and adaptation theories (which are themselves closely related). The framework is thus integrative in the sense that it nests all four elemental theories, but its principal function (at least in this section) is to differentiate among these theories by clarifying their distinctions.

Having emphasized in Section 4A this categorization of the theories into property-rights and incentive-systems on the one hand and adaptation and rent-seeking on the other, I then discuss in Section 4B alternative categorizations based on the distinctions among control, contract, and legal regimes.

4A. Ex Ante Incentives Versus Ex Post Governance

I begin by nesting just the property-rights and adaptation theories, via the following timing of information, actions, and payoffs (where, throughout this sub-section, in a slight generalization of the elemental theories, all the variables can be interpreted as vectors with arbitrary numbers of elements):

(1) Asset ownership / control of decision rights negotiated.

(2) Ex ante actions (a) chosen.

(3) State of the world (s) realized; state and actions publicly observed.

(4) Ex post decisions (d) taken.

(5) Payoffs $U_i(a, s, d)$ received.

Formally, what makes the actions in stage (2) "ex ante" but the decisions in stage (4) "ex post" is that the former are chosen before the state is realized but the latter are taken after. More concretely, it may help to interpret stage (2) as the production stage and stage (4) as the distribution stage (in the sense that in stage (4) the parties distribute the gains from production, although the distribution process need not be efficient).

In the property-rights theory, the ex ante actions are observable but not contractible, whereas the ex post decisions are contractible in stage (4), as required for the parties to bargain over these decisions in stage (4), but are not contractible in earlier stages. (Denote ex ante actions that are observable but not contractible by a^O and ex post decisions that are contractible ex post but not ex ante by d^P.) In the adaptation theory, in contrast, the ex ante actions do not exist, and the ex post decisions are not contractible in any stage. (Denote ex post decisions that are not contractible ex post by d^N.)

Comparing the ingredients of the elemental property-rights and adaptation theories shows that they are by no means identical; to the contrary, they are essentially orthogonal. To borrow a distinction from Williamson (2000), the property-rights theory emphasizes "ex ante incentive alignment," whereas the adaptation theory emphasizes "ex post decision governance" (two phrases that are roughly synonymous with what I mean by production and distribution, respectively). More specifically, in comparing transaction-cost economics (TCE) to the property-rights approach of Grossman-Hart-Moore (GHM), Williamson (2000: 605) 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 costless ex post bargaining. The upshot is that all of the inefficiency in GHM is concentrated in the ex ante investments in human assets (which are conditional on the ownership of physical assets)."

To nest the incentive-system and rent-seeking theories with the property-rights and adaptation theories in the framework begun above, I will apply Williamson's distinction between ex ante incentive alignment and ex post decision governance to all four theories. The result is shown in Table 2: the incentive-system and property-rights theories focus on ex ante incentive alignment, whereas the adaptation and rent-seeking theories focus on ex post decision governance.

Ex ante Incentive Alignment		Ex post Decision Governance	
Incentive Systems	Property Rights	Adaptation	Rent-Seeking
Holmstrom-Milgrom 91/94 Holmstrom-Tirole 91 Holmstrom 99	Grossman-Hart 86 Hart-Moore 90 Hart 95	Simon 51 Williamson 71/75/91 Klein-Murphy 88, 97 Klein 96, 00a	Williamson 71/79/85 Klein-Crawford-Alchian 78 Klein 88/00b

Table 2: Incentive Alignment vs. Decision Governance

Simply put, my view is that there are really two major theories of the firm, each with various special cases (perhaps not limited to the two special cases of each major theory shown here). The purpose of this sub-section is to make this two-theory argument formally, by enriching the integrative framework begun above.

To expand the framework to include the incentive-system theory with the property-rights and adaptation theories, we need to introduce the performance measure p and the

contract $w = s + bp$. The contract negotiation can be included in stage (1), the observation of p can be included in stage (3), and the payment of w can be included in stage (4). But even with these expansions of the framework, there are several aspects of the incentive-system theory that remain to be captured. First, in keeping with a traditional agency model, the ex ante actions in the incentive-system theory are not observable, whereas the ex ante actions in the property-rights theory are observable. (Denote ex ante actions that are not observable by a^N.) Second, the performance measure p in the incentive-system theory is contractible, whereas the state and the ex ante actions in the property-rights model are observable but not contractible. All of these variables can be interpreted as "interim signals" that are observed in stage (3), but some are contractible and others are not. (Denote interim signals that are contractible by σ^C and those that are observable but not contractible by σ^O.) Finally, the wage paid in the incentive-system model is a contractually obligated function of p, whereas the ex post decisions taken in the property-rights are contractible ex post but not ex ante. (Denote ex post decisions that are contractible ex ante by d^A.) These expansions of the integrative framework result in the following timing:

(1) Governance structure negotiated: control of d^P and d^N; contracts $d^A(\sigma^C)$

(2) Ex ante actions chosen: $a = (a^O, a^N)$

(3) Interim signals observed: $\sigma = (\sigma^C, \sigma^O)$

(4) Ex post decisions taken: $d = (d^A, d^P, d^N)$

(5) Payoffs $U_i(a, \sigma, d)$ received.

It remains to nest the rent-seeking theory into this framework. For this purpose, I will take the elemental rent-seeking theory to be the theory sketched in connection with (20), where the decision rights d_1 and d_2 are alienable (*i.e.*, can be controlled by either party), but the decision rights η_i and η_j are inalienable (*i.e.*, can be controlled only by parties i and j, respectively). To nest the rent-seeking theory into the integrative framework, we need to know more about the prize that inspires the rent-seeking (namely, AQR) than was discussed in connection with (20). In particular, we need to know something about the source of AQR. I will assume that AQR results from contractible ex ante actions, denoted by a^C. (For example, imagine that GM and Fisher Body wrote a perfect contract specifying exactly where a plant should be located and how it should be designed and built. Once the plant is built, precisely to specifications, there are appropriable quasi-rents to be had.) Thus, the new ingredients in the rent-seeking theory are: contractible ex ante actions (a^C); inalienable ex post decisions (η^N), which are not

contractible ex post; and, for completeness, perhaps new contracts that specify the ex post decisions that are contractible ex ante (d^A) in part as functions of these new contractible ex ante actions. In sum, we have:

Integrative Framework for Four Elemental Theories:

(1) Governance structure negotiated: control of d^P and d^N; contracts $d^A(a^C, \sigma^C)$

(2) Ex ante actions chosen: $a = (a^C, a^O, a^N)$

(3) Interim signals observed: $\sigma = (\sigma^C, \sigma^O)$

(4) Ex post decisions taken: $d = (d^A, d^P, d^N, \eta^N)$

(5) Payoffs $U_i(a, \sigma, d)$ received.

Having constructed this integrative framework, I now ask how it can be used. I see two purposes. First, one can use an integrative framework to differentiate among theories. The ingredients of the four elemental theories (as I have sketched them) are: property-rights = (a^O, σ^O, d^P), adaptation = (σ^O, d^N), incentive-system = (a^N, σ^C, d^A), and rent-seeking = (a^C, d^N, η^N). Thus, for example, the elemental rent-seeking and property-rights theories are orthogonal, in the sense that they have disjoint sets of ingredients.[12] Second, one can use this framework to address the question mark in this essay's title, by asking "How many theories of the firm are there, really?"

As we have seen at various points, the incentive-system and property-rights theories of the firm are closely related, in the sense that both focus on creating incentives for efficient ex ante actions to be chosen in stage (2). The incentive-system theory creates these incentives via a contract $w = s + bp$ negotiated in stage (1), with the performance measure p observed in stage (3) and the wage w paid in stage (4). In contrast, the property-rights theory has no incentive contracts, but the allocation of asset ownership affects the parties' threat points in negotiating over the ex post decisions that are contractible in stage (4), and hence the allocation of surplus ex post, and hence the parties' incentives for ex ante action choices in stage (2). These two theories are thus special cases of the following:

[12] The ingredients of the influence-cost theory in Section 3D are somewhat tricky to fit into this integrative framework. In effect, I have moved η^N to stage (2), so that the ingredients are η^N, σ^O, and d^N. But note that these inalienable influence actions in stage (2) differ from the production / ex ante incentive interpretation of the ex ante actions that we have otherwise maintained.

Framework for Models of Ex Ante Incentive Alignment:

(1) Governance structure negotiated: control of d^P; contracts $d^A(\sigma^C)$

(2) Ex ante actions chosen: $a = (a^O, a^N)$

(3) Interim signals observed: $\sigma = (\sigma^C, \sigma^O)$

(4) Ex post decisions taken: $d = (d^A, d^P)$

(5) Payoffs $U_i(a, \sigma, d)$ received.

As we have also seen at various points, the rent-seeking and adaptation theories of the firm are closely related, in the sense that both focus on creating incentives for efficient ex post decision governance in stage (4). The rent-seeking theory allocates alienable decision rights so as to minimize rent-seeking undertaken not only with these alienable decision rights but also with inalienable decision rights. In contrast, the adaptation theory has no inalienable decision rights but explicitly considers the way different allocations of alienable decision rights facilitate "adaptive, sequential decision-making" as uncertainty is resolved. These two theories are thus special cases of the following:

Framework for Models of Ex Post Decision Governance:

(1) Governance structure negotiated: control of d^N; contracts on a^C

(2) Ex ante actions chosen: $a = a^C$

(3) Interim signals observed: $\sigma = \sigma^O$

(4) Ex post decisions taken: $d = (d^N, \eta^N)$

(5) Payoffs $U_i(a, \sigma, d)$ received.

To summarize: my view is that there are really two major theories of the firm, each with various special cases. One can make this argument informally, as in Table 2. This sub-section made this argument formally, by developing an integrative framework that distinguishes not only between the major theories but also between some of their special cases.

4B. Alternative Categorizations

I find it productive to distinguish between ex ante incentive alignment versus ex post decision governance. But there is another way to divide the four elemental theories described in this essay: into the "control" versus "contract" branches of the theory of the

firm. I discuss this control-versus-contract distinction here – partly in the spirit of due diligence (*i.e.*, to complement my emphasis of the incentives-versus-governance distinction), but also because I think some of the future work described in the Conclusion may help clarify (or even erase) the control-versus-contract distinction.

The control branch of the theory of the firm asserts that integration gives greater control (over something). As noted in Section 1, this idea has a long history in industrial organization; for example, this is why an integrated supply chain avoids double marginalization. In the theory of the firm, the control branch dates from Coase's (1937: 388) observation that "Within a firm ... market transactions are [replaced by] the entrepreneur–coordinator, who *directs* production" (emphasis added). From its earliest days, the informal rent-seeking theory adopted Coase's approach; see, for example, Williamson's (1971: 114) discussion of how integration allows decisions to be made by fiat. There are also several formal theories in this control tradition: Simon's (1951) model of authority in the employment relationship seems to be the first, but the property-rights and adaptation theories described here are also squarely in this tradition.

In contrast, the contract branch denies that integration changes anything. In fact, this approach can be seen as denying that integration *is* anything (besides a label). Instead, the contract branch envisions a set of heterogeneous transactions and a fixed set of feasible contracts: different collections of transactions warrant different choices of contracts from this fixed feasible set, but the solution to each transaction is purely contractual (even if the contracts look and feel like something we might call a "firm"). That is, in the contract tradition, "the word *firm* is merely descriptive, a collective noun denoting a particular cluster of otherwise ordinary contractual relationships" (Masten, 1988: 181).

The contract tradition stems from informal theory originating principally at Chicago, Rochester, and UCLA.[13] If a tradition can originate from two sentences, in this case they are probably Alchian and Demsetz's (1972: 777) famously denial that integration changes anything: "It is common to see the firm characterized by the power to settle disputes by fiat, by authority, or by disciplinary action superior to that available in the conventional market. This is a delusion." In the Alchian-Demsetz account, team production causes monitoring difficulties that are best addressed by hiring a monitor who, alone, is the central party to all contracts with team members, can alter team

[13] For example, see Alchian and Demsetz (1972), Jensen and Meckling (1976), Cheung (1983), Fama and Jensen (1983), and Demsetz (1988).

membership, and is the residual claimant of the team's output (after payments to team members). Jensen and Meckling (1976: 310-11) endorsed the contracting spirit of Alchian-Demsetz, but argued that their focus on team production was too narrow:

> Contractual relations are the essence of the firm, not only with employees but with suppliers, customers, creditors, and so on. The problem of agency costs and monitoring exists for all of these contracts, independent of whether there is [team] production. ... [As a result], it makes little or no sense to try to distinguish those things that are 'inside' the firm (or any other organization) from those things that are 'outside' of it. There is in a very real sense only a multitude of complex relationships (*i.e.*, contracts) between the legal fiction (the firm) and the owners of labor, material and capital inputs and the consumers of output.

Finally, Cheung (1983: 16-18) gave a detailed analysis of a wide range of contractual forms, summarized by:

> I have traced the transitions from the simple service of a middleman, through various piece-rate arrangements, to wage contracts. The measurements adopted range respectively from a whole product to a component to a proxy measure. The information of the price signal moves from a full valuation of the contribution to no direct signal at all. Direction and monitoring by the [principal] correspondingly rise in complexity. Delegation of the right to use the input also increases until full control is granted in terms of some contractual limits. ... Thus it is futile to press the issue of what is or is not a firm. ... The important questions are why contracts take the forms observed and what are the economic implications of different contractual and pricing arrangements.

I find these ideas an enormously promising start to what could be called "contract economics," on which more below. So far, however, there are far fewer formal models in the contract branch of the theory of the firm than in the control branch. The elemental incentive-system theory presented here is in the contract spirit (because whether the agent owns the asset does not affect the agent's action space, so asset ownership does not reallocate control rights), but the richer incentive-system models by Holmstrom and Milgrom (1991, 1994), Holmstrom and Tirole (1991), and Holmstrom (1999) incorporate control considerations as well. More recently, formal models closer to the contract tradition have been developed by Bajari and Tadelis (2001), Tadelis (2002), and Levin and Tadelis (2004). Bajari and Tadelis compare two contract forms (fixed-price and cost-plus), but develop their model in the context of contracting between firms with fixed boundaries, leaving the analogy to the make-or-buy problem for their discussion section.

Tadelis takes the next step, explicitly modeling the make-or-buy decision, but with a reduced-form model that could be interpreted as being in the contract tradition, the control tradition, or a blend of the two. Finally, Levin and Tadelis develop the first formal model I know that is squarely in the contract tradition. There are heterogeneous transactions and many forms of contract, ranging from one that looks like employment to another that looks like outsourcing. In equilibrium, only these two extreme contract forms are chosen. Transactors who choose the former are called integrated, while those who choose the latter are called non-integrated, even though both governance structures are just contracts.

I raised this control-versus-contract distinction both in the spirit of due diligence and also because I think some of the future work described in the Conclusion may help clarify (or even erase) this distinction. But control-versus-contract is not the only alternative to the distinction between ex ante incentives versus ex post governance. In addition, one can distinguish between different legal regimes in which different contracts are feasible. For example, Masten (1988) compares employment law to commercial law and finds that an employer has control rights over an employee (*e.g.*, concerning the manner in which work is performed, as opposed to just the outcome of the work) that a firm does not have over an independent contractor. Similarly, Williamson (1991: 275) argues that:

> The implicit contract law of internal organization is that of forebearance. Thus, whereas courts routinely grant standing to firms should there be disputes over prices, the damages to be ascribed to delays, failures of quality, and the like, courts will refuse to hear disputes between one division and another over identical technical issues.

This Masten-Williamson distinction between legal regimes again differs from the four elemental theories emphasized in this essay, but in the opposite direction from the contracting tradition just described. Unlike the contracting tradition, these elemental theories (perhaps excepting the elemental incentive-system theory) assert that integration gives greater control over something. Unlike the legal-regimes arguments, however, these elemental theories are rooted in a single legal regime: integration does not change what contracts are feasible (or what information is observable, or what preferences the parties have, or anything else, except who controls something).

5. Conclusion

I hope to have presented clear words and models concerning four elemental theories of the firm. By having pulled these theories apart, I hope to encourage others to put them together, in new and interesting ways, in both theoretical and empirical work.

To conclude this essay, I begin with brief summaries of the four theories. I then describe two sets of issues that I think deserve attention in future work: first, issues that fit within the close confines set by the four caveats given in the Introduction ("pushing the paradigm"); second, issues that do not fit within these confines ("breaking loose").

5A. Four Elemental Theories of the Firm?

The distinctive point in (this essay's telling of) the incentive-system theory was that asset ownership can be an instrument in a multi-task incentive problem. In the elemental version of this theory, integration reallocates only payoff rights, not decision rights; the latter are constant in the agent's moral hazard problem. Reallocating payoff rights via changes in asset ownership can facilitate new incentive contracts that improve total incentives (arising from the contract and asset ownership).

The other three theories – adaptation, property-rights, and rent-seeking – all conceive of integration as reallocating decision rights, rather than payoff rights. Of these three theories, the adaptation theory is the simplest: control is valuable because it will be exercised. One important insight from (this telling of) the adaptation theory is that, by focusing on "adaptive, sequential decision-making," we can develop a theory of the firm without specific investments.

The property-rights theory of the firm is more complex than the incentive-system theory or the adaptation theory: it emphasizes how integration reallocates decision rights, as in the adaptation theory, but the efficiency consequences of these reallocated decision rights appear in ex ante actions, akin to the incentive-system theory. One insight from (this telling of) the property-rights theory is that using a formal instrument to stop one hold-up problem typically creates another hold-up problem. A second insight (a corollary of the first, but cast more concretely) is that the cost of control is the loss of initiative.

Finally, the rent-seeking theory gave early and sharp insights into the transaction costs of non-integration, such as the hypothesis that the inefficiency from haggling increases with the AQRs. More formally, a distinctive point in (this telling of) the rent-

seeking theory of the firm is that ownership can stop haggling that is undertaken via alienable instruments. But, of all four theories, the rent-seeking theory occupies the strangest position: it has been hardest to formalize, in the sense of providing a unified account of the costs and benefits of integration, yet it has launched by far the most empirical work.

5B. Future Work, I: Pushing the Paradigm

Even within the close confines set by the four caveats given in the Introduction (concerning scope, coverage, methodology, and style), there is much important work to be done that deepens and applies the four theories described here. I think the following three areas are especially promising: contracting for control, hybrid governance structures, and relational contracting.

Contracting for control: As noted in Section 2, the adaptation theory applies not only to the make-or-buy problem but also (and at least as well) to a particular class of contracting problems, where two firms with fixed boundaries pass decision rights across their boundaries by contract. Maskin and Tirole (1999) make a parallel point when they note the formal equivalence between the property-rights theory and certain contracting models. In short, the make-or-buy theories described here may have as much to say about contracts *between* firms as they do about the boundaries *of* firms.

Turning from theory to evidence, there has recently been a surge of empirical work in "contract economics." After early contributions such as Goldberg and Erickson (1987), Joskow (1985, 1987), Masten and Crocker (1985), and Palay (1984), there was something of a lull in this literature (with a few conspicuous exceptions, such as the work of Lafontaine and co-authors[14]). More recently, however, contract economics has sprung back to life; see, for example, the 26 essays in Brousseau and Glachant (2002). In particular, many recent empirical papers can be seen as analyzing issues akin to contractual movements of decision rights across fixed firm boundaries – including Arruñada, Garicano, and Vazquez (2001), Bidwell (2004), Elfenbein and Lerner (2003), Kaplan and Stromberg (2003), Lafontaine and Masten (2002), Lerner and Merges (1998), and Robinson and Stuart (2002), who analyze car dealerships, software development, internet portals, venture capital, trucking, and biotechnology.

---

[14] For example, see Lafontaine (1992, 1993), Bhattacharya and Lafontaine (1995), Lafontaine and Slade (1997), and Lafontaine and Shaw (1999).

Several next steps would be useful. First, it would be a good start to have a theory of passing decision rights across fixed firm boundaries: taking firm boundaries as given, which decision rights should be transferred in which circumstances? Second, the firm boundaries should be endogenized: should a decision right be reallocated by contract, or should an entire asset be reallocated by ownership? Third, this line of research seems likely to shed light on (and perhaps even erase) the contract-versus-control distinction discussed in Section 4B, as follows.

As one way to model the control-versus-contract distinction (and the "contracting for control" empirical literature just described), I will follow Baker, Gibbons, and Murphy (2004a) by assuming that an asset consist of three components: a vector of extricable decision rights, d (where extricable means that these decision rights can be moved by contract, without changing asset ownership); a vector of inextricable decision rights, δ (where inextricable means that these decision rights are controlled by the asset owner); and an inextricable payoff right, π (where inextricable now means that this payoff is received by the asset owner). In this environment, "contracting for control" means moving control of (all or part of) d from the asset's owner to another party, whereas "integration" (or any other restructuring of the firms' boundaries) means reallocating the asset's ownership from the current owner to another party.

I will continue to follow Baker, Gibbons, and Murphy (2004a) by assuming that d, δ, and π are not contractible ex post. Two special cases of this model then should seem familiar. First, if only π exists (*i.e.*, d and δ are immaterial), then this model is akin to the elemental incentive-system model from Section 2A (although I have said nothing about the ex ante actions that appear in that model). Second, if only d exists (*i.e.*, δ and π are immaterial), then this model is akin to the elemental adaptation model from Section 2B (although I have said nothing about the state variable that appears in that model).

In this environment, one can express the central assertion of the contract branch of the theory of the firm – namely, that a firm is merely "a collective noun denoting a particular cluster of otherwise ordinary contractual relationships" – as the assumption that δ and π are immaterial. This assumption is sufficient (and essentially necessary) for each feasible governance structure to be equivalent to a contractual allocation of extricable decision rights; that is, under this assumption, there is nothing special about "ownership," "integration," or "firms." It is then an empirical question whether this assumption is a useful approximation to reality often, sometimes, or rarely. We should let the data speak.

Hybrid governance structures:[15] Coase (1937: 388) famously quoted Robertson's description of firms as "islands of conscious power ... like lumps of butter coagulating in a pail of buttermilk." Notwithstanding the "butter" part of this metaphor, the "islands" part suggests that firm boundaries are sharp, in the sense that one firm is clearly separated from another. All four of the theories reviewed here have such sharp boundaries of the firm.

While this "islands" view has been productive both theoretically and empirically, various dissenting and complementary views have occasionally surfaced. Even in 1937, Coase cautioned that "it is not possible to draw a hard and fast line which determines whether there is a firm or not "(p. 392). There ensued both theoretical discussion about *what a firm is* and empirical documentation of *what exists besides firms*. Having indulged in some of the former in the control-versus-contract discussion above, I now turn to the latter. Richardson (1972) is a classic in this vein, giving a rich description of "industrial activity that our simple story, based as it is on a dichotomy between firm and market, leaves out of account ... [namely,] the dense network of co-operation and affiliation by which firms are inter-related." Cheung's (1983) description of many contractual structures between firm and market, Eccles's (1981) work on quasi-firms, and Powell's (1990) discussion of networks ("Neither Market Nor Hierarchy") all continue this tradition of empirically grounded criticism of the simple dichotomy between markets and firms.

The many governance structures besides firms and markets are sometimes summarized as "hybrid" governance structures; see Williamson (1985; 1996) and Menard (2004). Contracts that pass decision rights across fixed firm boundaries can be seen as one form of hybrid, but there are many other forms. In fact, even brief inspection of the existing governance structures in industries such as pharmaceuticals, biotechnology, medical devices, airlines, and telecommunications shows that firms have invented far more ways to work together than organizational economics has so far expressed (not to mention evaluated). Much work remains to be done; see Baker, Gibbons, and Murphy (2004b) for a start and further references.

Relational contracts:[16] As noted in Section 1, all the theories reviewed here have recognized the crucial role of contract imperfections in making integration matter.

[15] This discussion draws heavily on Baker, Gibbons, and Murphy (2004b).

[16] This discussion draws heavily on Baker, Gibbons, and Murphy (2002), which also provides many additional references.

Indeed, in these terms, the shared message of all four theories can be stated as follows: if contract imperfections are wreaking havoc under one governance structure, consider changing to another (*e.g.*, integrate a transaction formerly conducted under non-integration). As hinted in Section 1D, however, there is another possible remedy when contracts are imperfect: leave the governance structure alone, but move to "relational contracting."

A relational contract is a self-enforcing agreement so rooted in the parties' particular circumstances that the agreement cannot be enforced by a third party, such as a court. Relational contracts may circumvent difficulties in formal contracting (*i.e.*, contracting enforced by a court). For example, a formal contract must be specified ex ante in terms that can be verified ex post by the third party, whereas a relational contract can be based on outcomes that are observed ex post by only the contracting parties and also on outcomes that are prohibitively costly to specify ex ante. A relational contract thus allows the parties to utilize their detailed knowledge of their specific situation and to adapt to new information as it becomes available. For the same reasons, however, relational contracts cannot be enforced by a third party and so must be self-enforcing: the value of the future relationship must be sufficiently large that neither party wishes to renege. Accordingly, contracts here called "relational" are sometimes called "self-enforcing" (Klein and Leffler, 1981; Telser, 1981), "implicit" (MacLeod and Malcomson, 1989, 1998), or both (Bull, 1987). A common approach to modeling such relational contracts is to use a repeated game, just as models of tacit collusion in industrial organization have used repeated games for several decades; see Levin (2003) for the state of the art.

Sociologists (and other non-economists with substantial experience inside organizations) have argued for decades that firms are riddled with relational contracts – informal agreements and unwritten codes of conduct that powerfully affect the behaviors of individuals within firms. There are often informal quid pro quos between co-workers, as well as unwritten understandings between bosses and subordinates about task-assignment, promotion, and termination decisions. Even ostensibly formal processes such as compensation, transfer pricing, internal auditing, and capital budgeting often cannot be understood without consideration of their associated informal agreements.

Furthermore, business dealings are also riddled with relational contracts. Supply chains often involve long-run, hand-in-glove supplier relationships through which the parties reach accommodations when unforeseen or uncontracted-for events occur. Similar

relationships also exist horizontally, as in the networks of firms in the fashion industry or the diamond trade, and in strategic alliances, joint ventures, and business groups. Whether vertical or horizontal, these relational contracts influence the behaviors of firms in their dealings with other firms.

Theoretically, we have known since at least Friedman (1971) that the shadow of the future can cause parties to behave in ways that are collectively useful but not in their short-run individual interest. So it is not surprising that relational contracts *can* help parties remedy imperfect formal contracts (a theoretical statement), and it is only mildly interesting that relational contracts *do* help parties remedy imperfect formal contracts (an empirical statement). What is important, however, is that the formal and the informal interact. In the language of economic theory, choosing a different governance structure can change the set of feasible relational contracts. In the language of sociology (Blau and Scott, 1962: 6):

> It is impossible to understand the nature of a formal organization without investigating the networks of informal relations and the unofficial norms as well as the formal hierarchy of authority and the official body of rules, since the formally instituted and the informal emerging patterns are inextricably intertwined.

Baker, Gibbons, and Murphy (1994, 1999, 2001, 2002, 2004a, 2004b) develop models that explore this interaction between formal and informal. The specific applications and conclusions differ across settings – compensation, delegation, supply chains, contracts, and alliances – but the overarching themes of this work are that (a) relational contracts can and do help parties remedy imperfect formal governance structures (including but not limited to imperfect formal contracts), and (b) formal and informal interact, so (c) the formal governance structure should be chosen not only for its own impacts but also for how it affects the feasible set of relational contracts.

As suggested in the Section 1D (in the context of the adaptation theory), Klein and Williamson emphasized opposite halves of a story along these lines: while Klein focused on relational contracting between firms (which Williamson downplayed), Klein downplayed relational contracting within firms (which Williamson accentuated). Models like those of Baker, Gibbons, and Murphy show that Klein's and Williamson's approaches can be unified not only in the adaptation theory, but in each of the elemental theories described here. Whatever the elemental theory, the question is whether integration or non-integration facilitates superior relational contracting. The key point in

the resulting relational-contracting theory of the firm is then that the integration decision is chosen to facilitate the parties' relationship.

5C. Future Work, II: Breaking Loose

I placed several limits on this essay. First, I limited the scope of the essay by focusing on "the theory of the firm" (*i.e.*, Coase's make-or-buy problem), with scant attention paid to related topics in organizational economics (such as structures and processes inside organizations). I think it is important to recall the broader usage intended by Cyert and March (1963), for whom "theory of the firm" meant descriptive and prescriptive models of firms' decision-making processes. There is now a great deal of work consistent with this broader usage, but much of this work focuses on an individual transaction (albeit one inside a firm) and so retains the spirit of the theories of the firm discussed here. Winter (1988, 2001) makes a persuasive case that this focus on "exchange" (whether between firms or within) has short-changed an equally important research agenda on "production," which should analyze a host of issues that are swept under the rug by the standard production-function approach, such as "replication" (*i.e.*, whether a firm can repeat tomorrow what it did successfully today). Ultimately, these two research agendas – production and exchange – should not only co-exist but interact; see Jacobides and Winter (2003), Langlois (1992), and Langlois and Foss (1999) for more along this line.

Second, I limited the coverage of this essay by focusing on just four theories, almost entirely ignoring a wide variety of others – formal, informal, and gestating. Additional formal theories include Matouschek (2004), McAfee and McMillan (1995), Rajan and Zingales (1998), Rotemberg and Saloner (1994), Tadelis (2002), and Wernerfelt (1997). But all of these theories concern incentives, in one form or another. Indeed, a huge share of the literature bears a family resemblance to the four theories discussed here via this emphasis on (ex ante or ex post) incentives. An alternative approach could explore the complexity of coordination and the limits that bounded rationality consequently places on firm size and scope. These difficulties of complex coordination may be closer to Coase's (1937) original ideas than are any of the theories described here: "in place of the complicated market structure with exchange transactions is substituted the entrepreneur ... who directs production" (p. 388); "as the transactions which are organized increase, the entrepreneur fails to place the factors of production in

the uses where their value is greatest" (p. 394). See Garicano (2002), Van Zandt (1998), and Williamson (1967) for starts in this direction.

Third, I limited the methodology of this essay by discussing theory, essentially ignoring evidence and testing. Nonetheless, I hope that this essay facilitates new applications of Whinston's (2003) meta-analytic approach: modern theories can be evaluated by revisiting older evidence, especially by looking for aspects of the evidence that the modern theories highlight but the original investigators had no reason to emphasize. More generally, I hope that by providing elemental versions of these four theories, both those who revisit older evidence and those who gather new evidence will have a richer sense of what the various theories say, and hence what to look for in the data. But we should not expect a horserace among these four elemental theories, for three reasons. First, these four theories are certainly not exclusive: they could all be true; see Azoulay (2004) for evidence in this spirit. Second, these elemental theories are quite abstract, so empirical researchers may need to tailor models to their empirical domains; see Baker and Hubbard (2003) for theory and evidence in this mold. Finally, it remains unclear (at least to me) whether the large existing empirical literature relating asset specificity to vertical integration can be seen as testing an elemental rent-seeking theory of the firm (and, if so, what that elemental rent-seeking theory might be).

Fourth, and perhaps most important, I limited the style of this essay by presenting stick-figure introductions, not full-bodied statements of the original theories. Regarding the two theories that were formal from their beginnings (property-rights and incentive-system), this limitation bothers me somewhat, but interested readers can guide themselves from the introductions given here to richer formal statements in the literature. Regarding the two theories that were originally informal (rent-seeking and adaptation), however, this limitation bothers me greatly, for two reasons: first, the stick-figures I presented may have missed the main points of these informal arguments; second, and more important, even if I captured the main points, the distillation of this large body of work into two stick-figure models risks greatly understating the crucial role that this work played in developing the field. I conclude this essay by unpacking these two concerns.

Many economists seem to take it for granted that formalization is good. To a large extent, I agree. In particular, I think there are at least three reasons why it is important to try to formalize informal arguments (Gibbons, 1999: 152-4). First, formal models check

the internal consistency of informal arguments.[17] Second, formal models can help to specify and interpret empirical tests.[18] Third, developing formal models in rich contexts sometimes allows the analysis to push through to conclusions when informal analysis would (or should) have ground to a halt.[19] In my view, Section 3's discussion of the rent-seeking theory of the firm illustrates all three of these potential virtues of formalization.

On the other hand, formalization can have costs. For example, Kreps (1996: 562) argues that "If *Markets and Hierarchies* has been translated into game theory using notions of information economics, it is a very poor translation. ... Anyone who relies on the translations alone misses large and valuable chunks of the original." Similarly, Williamson (2000: 604-5) argues that "Formalization is vital to a progressive research agenda, but ... provision also needs to be made for the possibility that core features of the theory are left out or obscured by the translation." Williamson then uses the distinction I borrowed in Section 4 – between "ex ante incentive alignment" and "ex post decision governance" – to argue that the formal property-rights theory has obscured a key idea from the informal rent-seeking theory.

From these costs and benefits of formalization, I conclude that we want literatures to blend detailed description, informal theory, and formal modeling (as well as persuasive econometrics and experiments, but I am discussing theory here). Unfortunately, I think it is far too easy to understate the value of not only informal theory but especially detailed description, as illustrated by the following confessions of an enthusiastic model-builder.

When I first read Coase's (1984: 230) description of the collected works of the old-school institutionalists – as "a mass of descriptive material waiting for a theory, or a fire" – I thought it was (a) hysterically funny and (b) surely dead-on (even though I had not

---

[17] For example, see Holmstrom's (1982) formalization of Fama's (1980) informal argument that managerial incentive contracts are not necessary because managers are disciplined career concerns (*i.e.*, by the way that a manager's future value in the labor market responds to her current performance).

[18] For example, in attempting to test Spence's (1973) signaling model of education, Layard and Psacharopoulos (1974) proposed a simple and appealing test that seemed to be implied by an unmodeled dynamic extension of Spence's theory, but Farber and Gibbons (1996) developed a formal model of these dynamics and showed that this seemingly intuitive but informally deduced empirical test was not in fact implied by the theory in question.

[19] The sociologist James Coleman (1964: vii) made this point much more eloquently: "If conceptual elaboration is to progress beyond the proverbs of the ancients, special tools are necessary. The most remarkable of these is mathematics. ... The mind falters when faced with a complex system or a long chain of deductions. The crutch that mathematics provides to everyday reasoning becomes essential as sociology moves toward the analysis of complex systems and predictions based on extended chains of deductions."

read this work). Sometime later, I encountered Krugman's (1995: 27) assertion that "Like it or not, ... the influence of ideas that have not been embalmed in models soon decays." I think my reaction to Krugman was almost as enthusiastic as my reaction to Coase, although I hope the word "embalmed" gave me at least some pause. But then I made it to Krugman's contention that a prominent model in economic geography "was the one piece of a heterodox framework that could easily be handled with orthodox methods, and so it attracted research out of all proportion to its considerable merits" (p. 54). At this point, I stopped reading and started trying to think.

In my view, we are stuck with "Like it *and* not." That is, formalization has huge benefits (which do not need support or even articulation from the likes of me). But we must avoid Krugman's "tales of frustration: sensible ideas that could not be effectively formalized, [and] formalizable ideas that seem to have missed the point" (p. 59). And Krugman's warning of course does not apply solely to economic geography. To the contrary, in work directly related to this essay, Goldberg and Erickson (1987: 369) make a similar point when they describe their case study of contracts for petroleum coke as exploring "the way in which reasonably clever businessmen and lawyers cope with problems scholars might consider intractable."

I hope this essay has managed to avoid "formalizable ideas that seem to have missed the point," but I hope this literature refuses to deem intractable various problems that "reasonably clever businessmen and lawyers cope with" routinely. In trying to balance these twin concerns, I have been enormously influenced by the detailed descriptions and informal theories from Klein and Williamson. I therefore end this essay as the Introduction ended: without three decades of work by these two tireless contributors, I expect that we would have fewer and narrower formal theories, and I am sure that we would be less well positioned to push our current and prospective formal theories towards a thorough understanding of organization and governance.

References

Aghion, Philippe, Mathias Dewatripont, and Patrick Rey. 2002. "Transferable Control." Unpublished manuscript.

Aghion, Philippe and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105:1-29.

Arruñada, Benito, Luis Garicano, and Luis Vázquez. 2001. "Contractual Allocation of Decision Rights and Incentives: The Case of Automobile Distribution." *Journal of Law, Economics, and Organization* 17: 257-84.

Alchian, Armen and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62: 777-95.

Anderson, Erin and David Schmittlein. 1984. "Integration of the Sales Force: An Empirical Examination." *Rand Journal of Economics* 15: 385-95.

Azoulay, Pierre. 2004. "Capturing Knowledge Within and Across Firm Boundaries: Evidence from Clinical Development." Forthcoming in *American Economic Review.*

Bajari, Patrick and Steven Tadelis. 2001. "Incentives versus Transaction Costs: A Theory of Procurement Contracts." *Rand Journal of Economics* 32: 387-407.

Baker, George. 2002. "Distortion and Risk in Optimal Incentive Contracts." *Journal of Human Resources* 37: 728-51.

_____, Robert Gibbons, and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics* 109: 1125-56.

_____, _____, and _____. 1999. "Informal Authority in Organizations." *Journal of Law, Economics, and Organization* 15: 56-73.

_____, _____, and _____. 2001. "Bringing the Market Inside the Firm?" *American Economic Review Papers and Proceedings* 91: 212-18.

_____, _____, and _____. 2002. "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics* 117: 39-83.

_____, _____, and _____. 2004a. "Contracting for Control." Unpublished manuscript.

_____, _____, and _____. 2004b. "Strategic Alliances: Bridges Between 'Islands of Conscious Power.'" Unpublished manuscript.

_____ and Thomas Hubbard. 2003. "Make Versus Buy in Trucking: Asset Ownership, Job Design, and Information." *American Economic Review* 93: 551-72.

Bhattacharyya, Sugato and Francine Lafontaine. 1995. "Double-Sided Moral Hazard and the Nature of Share Contracts." *Rand Journal of Economics* 26: 761-81.

Bidwell, Matthew. 2004. "What Do Firms Do Differently? Comparing the Governance of Internal and Outsourced IT Projects." Unpublished manuscript, MIT's Sloan School.

Blau, Peter, and Richard Scott. 1962. *Formal Organizations: A Comparative Approach.* San Francisco, CA: Chandler Publishing.

Boerner, C. and Jeffrey Macher. 2002. "Transaction Cost Economics: An Assessment of Empirical Research in the Social Sciences." Unpublished manuscript.

Brousseau, Eric and Jean-Michel Glachant. 2002. *The Economics of Contracts: Theories and Applications*. Cambridge: Cambridge University Press.

Buchanan, James, Robert Tollison, and Gordon Tullock. 1980. *Toward a Theory of the Rent-Seeking Society*. College Station, TX: Texas A&M University Press.

Bull, Clive. 1987. "The Existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics* 102:147-59.

Casadesus-Masanell, Ramon and Daniel Spulber. 2000. "The Fable of Fisher Body." *Journal of Law and Economics* 43: 67-104.

Cheung, Steven. 1983. "The Contractual Nature of the Firm." *Journal of Law and Economics* XXVI: 1-21.

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

_____. 1984. "The New Institutional Economics." *Journal of Institutional and Theoretical Economics* 140: 229-31.

_____. 2000. "The Acquisition of Fisher Body by General Motors." *Journal of Law and Economics* 43: 15-31.

Coleman, James. 1964. *Introduction to Mathematical Sociology*. London: Collier-Macmillan Limited.

Cyert, Richard and James March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall. Second edition, 1992. Oxford: Blackwell Publishers.

Datar, Srikant, Susan Kulp, and Richard Lambert. 2001. "Balancing Performance Measures." *Journal of Accounting Research* 39: 75-92

Demsetz, Harold. 1988. "The Theory of the Firm Revisited." *Journal of Law, Economics, and Organization* 4: 141-63.

Eccles, Robert. 1981. "The Quasifirm in the Construction Industry." *Journal of Economic Behavior and Organization* 2: 335-57.

Elfenbein, Daniel, and Josh Lerner. 2003. "Ownership and control rights in Internet portal alliances, 1995-1999." *RAND Journal of Economics* 34: 356-69.

Fama, Eugene. 1980. "Agency Problems and the Theory of the Firm." *Journal of Political Economy* 88: 288-307.

_____ and Michael Jensen. 1983. "Separation of Ownership and Control." *Journal of Law and Economics* XXVI: 301-25.

Farber, Henry and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Quarterly Journal of Economics* 111: 1007-47.

Feltham, Gerald and Jim Xie. 1994. "Performance Measure Congruity and Diverstiy in Multi-Task Principal/Agent Relations." *The Accounting Review* 69: 429-53.

Freeland, Robert. 2000. "Creating Holdup Through Vertical Integration: Fisher Body Revisited." *Journal of Law and Economics* 43: 33-66.

Friedman, James. 1971. "A Non-cooperative Equilibrium for Supergames." *Review of Economic Studies* 38: 1-12.

Fudenberg, Drew and Jean Tirole. 1986. "A 'Signal-Jamming' Theory of Predation." *Rand Journal of Economics* 17: 366-76.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108: 874-904.

Gibbons, Robert..1999. "Taking Coase Seriously." *Administrative Science Quarterly* 44: 145-57.

_____. 2003. "Team Theory, Garbage Cans, and Real Organizations: Some History and Prospects of Economic Research on Decision-Making in Organizations." *Industrial and Corporate Change* 12: 753-87.

_____. 2004. "Incentives Between Firms (and Within)." Forthcoming in *Management Science*.

Goldberg, Victor and John Erickson. 1987. "Quantity and Price Adjustment in Long-Term Contracts: A Case Study of Petroleum Coke." *Journal of Law and Economics* XXX: 369-98.

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

Grout, Paul. 1984. "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach." *Econometrica* 52: 449-60.

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

_____ and Bengt Holmstrom. 2002. "Vision and Firm Scope." Unpublished manuscript.

_____ and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98: 1119-58.

Helper, Susan, John Paul MacDuffie, and Charles Sabel. 2000. "Pragmatic Collaborations: Advancing Knowledge While Controlling Opportunism." *Industrial and Corporate Change* 9: 443-88.

Henderson, Rebecca and Kim Clark. 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms." *Administrative Science Quarterly* 35: 9-30.

Holmstrom, Bengt. 1982. "Managerial Incentive Problems—A Dynamic Perspective." Republished in *Review of Economic Studies* 66 (1999): 169-82.

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

_____ and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55: 303-28.

_____ and _____. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7:24-52.

_____ and _____. 1994. "The Firm as an Incentive System." *American Economic Review* 84: 972-91.

_____ and John Roberts. 1998. "The Boundaries of the Firm Revisited." *Journal of Economic Perspectives* 12: 73-94.

_____ and Jean Tirole. 1989. "The Theory of the Firm." Chapter 2 in R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization*. Amsterdam: North-Holland.

_____ and _____. 1991. "Transfer Pricing and Organizational Form." *Journal of Law, Economics, and Organization* 7:201-28.

Jacobides, Michael and Sidney Winter. 2003. "Capabilities, Transaction Costs, and Evolution: Understanding the Institutional Structure of Production." Unpublished manuscript.

Jensen, Michael and Wiiliam Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3: 305-60.

Joskow, Paul. 1985. "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generation Plants." *Journal of Law, Economics, and Organization* 1: 33-80.

_____. 1987. "Contract Duration and Relationship-Specific Investment: Empirical Evidence from Coal Markets." *American Economic Review* 77:168-85.

Kaplan, Steven and Per Strömberg. 2003. "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts." *Review of Economic Studies* 70: 281-315.

Klein, Benjamin. 1988. "Vertical Integration as Organizational Ownership: The Fisher Body-General Motors Relationship Revisited." *Journal of Law, Economics, and Organization* 4: 199-213.

_____. 1996. "Why Hold-ups Occur: The Self-Enforcing Range of Contractual Relationships." *Economic Inquiry* 34: 444-63.

_____. 2000a. "The Role of Incomplete Contracts in Self-Enforcing Relationships." *Revue D'Economie Industrielle* 92: 67-80.

_____. 2000b. "Fisher-General Motors and the Nature of the Firm." *Journal of Law and Economics* 43: 105-41.

_____. 2004. "Fisher body-General Motors Once Again: What Is a Holdup?" Unpublished manuscript.

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

_____ and Keith Leffler. 1981. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy* 89: 615-641.

_____ and Kevin M. Murphy. 1988. "Vertical Restraints as Contract Enforcement Mechanisms." *Journal of Law and Economics* 31: 265-97.

_____ and _____. 1997. "Vertical Integration as a Self-Enforcing Contractual Arrangement." *American Economic Review* 87: 415-20.

Knight, Frank. 1921. *Risk, Uncertainty, and Profit.* New York: Hart, Schaffner, and Marx; repr. (1965) New York: Harper and Row.

Kogut, Bruce and Udo Zander. 1992. "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology." *Organization Science* 3: 383-97.

Kreps, David. 1996. "Markets and Hierarchies and (Mathematical) Economic Theory." *Industrial and Corporate Change* 5: 561-96.

Krugman, Paul. 1995. *Development, Geography, and Economic Theory*. Cambridge, MA: MIT Press.

Lafontaine, Francine. 1992. "Agency Theory and Franchising: Some Empirical Results." *Rand Journal of Economics* 23: 263-83.

_____. 1993. "Contractual Arrangements as Signaling Devices: Evidence from Franchising." *Journal of Law, Economics, and Organizations* 9: 256-89.

_____ and Scott Masten. 2002. "Contracting in the Absence of Specific Investments and Moral Hazard: Understanding Carrier-Driver Relations in US Trucking." Unpublished manuscript.

_____ and Kathryn Shaw. 1999. "The Dynamics of Franchise Contracting: Evidence from Panel Data." *Journal of Political Economy* 107: 1041-80.

_____ and Margaret Slade. 1997. "Retail Contracting: Theory and Practice." *Journal of Industrial Economics* 45: 1025.

Langlois, Richard. 1992. "Transaction-cost Economics in Real Time." *Industrial and Corporate Change* 1: 99-127.

_____ and Nicolai Foss. 1999. "Capabilities and Governance: The Rebirth of Production in the Theory of Economic Organization." *Kyklos* 52: 201-18.

Layard, Richard and G. Psacharopoulos. 1974. "The Screening Hypothesis and the Returns to Education." *Journal of Political Economy* 82: 985-98.

Lerner, Josh and Robert Merges. 1998. "The Control of Technology Alliances: An Empirical Analysis of the Biotechnology Industry." *Journal of Industrial Economics* 46: 125-56.

Levin, Jonathan. 2003. "Relational Incentive Contracts." *American Economic Review* 93: 835-57.

_____ and Steven Tadelis. 2004. "Employment versus Contracting in Procurement: Theory and Evidence from U.S. Cities." Unpublished manuscript.

MacLeod, Bentley, and James Malcomson. 1989. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica* 57: 447-80.

_____ and Daniel Parent. 1999. "Job Characteristics and the Form of Compensation." *Research in Labor Economics* 18: 177-242.

March, James. 1962. "The Business Firm as a Political Coalition." *Journal of Politics* 24:662-78.

Maskin, Eric and Jean Tirole. 1999. "Two Remarks on Property Rights." *Review of Economic Studies* 66: 139-49.

Masten, Scott. 1982. *Transaction Costs, Institutional Choice and the Theory of the Firm*. Ph.D. Dissertation, University of Pennsylvania.

_____. 1984. "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics* 27: 403-17.

_____. 1986. "Institutional Choice and the Organization of Production: The Make-or-Buy Decision." *Journal of Institutional and Theoretical Economics* 142: 493-509.

_____. 1988. "A Legal Basis for the Firm." *Journal of Law, Economics, and Organization* 4: 181-98.

_____ and Keith Crocker. 1985. "Efficient Adaptation in Long-Term Contracts: Take-or-Pay Provisions for Natural Gas." *American Economic Review* 75: 1083-93.

_____, James Meehan, and Edward Snyder. 1991. "The Costs of Organization." *Journal of Law, Economics, and Organization* 7: 1-25.

Matouschek, Niko. 2004. "Ex Post Inefficiencies in a Property Rights Theory of the Firm." *Journal of Law, Economics, and Organization* 20: 125-47.

McAfee, Preston and John McMillan. 1995. "Organizational Diseconomies of Scale." *Journal of Economics and Management Strategy* 4: 399-426.

Ménard, Claude. 2004. "The Economics of Hybrid Organizations." Forthcoming in *Journal of Institutional and Theoretical Economics*.

Meyer, Margaret, Paul Milgrom, and John Roberts. 1992. "Organizational Prospects, Influence Costs, and Ownership Changes." *Journal of Economics and Management Strategy* 1: 9-35.

Milgrom, Paul and John Roberts. 1988. "An Economic Approach to Influence Activities in Organizations." *American Journal of Sociology* 94: S154-79.

_____ and _____. 1990. "Bargaining costs, influence costs, and the organization of economic activity." Chapter 3 in J. Alt and K. Shepsle (eds.), *Perspectives on Positive Political Economy*. Cambridge: Cambridge University Press.

Monteverde, Kirk, and David Teece. 1982. "Supplier Switching Costs and Vertical Integration in the Automobile Industry." *Bell Journal of Economics* 13: 206-12.

Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Nonaka, Ikujiro, and Hirotaka Takeuchi. 1995. *The Knowledge-Creating Company*. New York: Oxford University Press.

Palay, Thomas. 1984. "Comparative Institutional Economics: The Governance of Rail Freight Contracting." *Journal of Legal Studies* 13:265-87.

Penrose, Edith. 1959. *The Theory of the Growth of the Firm*. New York: Wiley.

Perez-Castrillo, J. and T. Verdier. 1992. "A General Analysis of Rent-Seeking Games." *Public Choice* 73: 335-50.

Pfeffer, Jeffrey. 1981. "Sources of Power in Organizations." Chapter 4 in *Power in Organizations*. Cambridge, MA: Ballinger.

Powell, Walter. 1990. "Neither Market Nor Hierarchy: Network Forms of Organization." *Research in Organizational Behavior* 12:295-336.

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

_____ and _____. 2000. "The Tyranny of Inequality." *Journal of Public Economics* 76: 521-58.

Richardson, George. 1972. "The Organisation of Industry." *Economic Journal* 82: 883-96.

Riordan, Michael and Oliver Williamson. 1985. "Asset Specificity and Economic Organization." *International Journal of Industrial Organization* 3: 365-78.

Robinson, David, and Toby Stuart. 2002. "Just How Incomplete *Are* Incomplete Contracts? Evidence from Biotech Strategic Alliances." Unpublished manuscript.

Rotemberg, Julio and Garth Saloner. 1994. "Benefits of Narrow Business Strategies." *American Economic Review* 84: 1330-49.

Simon, Herbert. 1951. "A formal theory of the employment relationship." *Econometrica* 19: 293-305.

Skaperdas, Stergios. 1992. "Cooperation, conflict, and power in the absence of property rights." *American Economic Review* 82: 720-739.

Spence, A. Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87: 355-74.

Szidarovszky, Ferenc and Koji Okuguchi. 1997. "On the Existence and Uniqueness of Pure Nash Equilibrium in Rent-Seeking Games." *Games and Economic Behavior* 18: 135-40.

Tadelis, Steven. 2002. "Complexity, Flexibility, and the Make-or-Buy Decision." *American Economic Review* 92: 433-37.

Telser, Lester. 1981. "A Theory of Self-Enforcing Agreements." *Journal of Business*. 53: 27-44.

Tullock, Gordon. 1980. "Efficient Rent Seeking." Chpater 6 in *Toward a Theory of the Rent-Seeking Society*. J. Buchanan, R. Tollison, and G. Tullock (eds.), College Station, TX: Texas A&M University Press.

Van Zandt, Tim. 1998. "Real-time decentralized information processing as a model of organizations with boundedly rational agents." *Review of Economic Studies* 66: 633-58.

Wernerfelt, Birger. 1984. "A Resource-based View of the Firm." *Strategic Management Journal* 5: 171-80.

_____. 1997. "On the Nature and Scope of the Firm: An Adjustment-Cost Theory." *Journal of Business* 70: 489-514.

Whinston, Michael. 2003. "On the Transaction Cost Determinants of Vertical Integration." *Journal of Law, Economics, and Organization* 19: 1-23.

Williamson, Oliver. 1967. "Hierarchical Control and Optimum Firm Size." *Journal of Political Economy* 75: 123-38.

_____. 1971. "The Vertical Integration of Production: Market Failure Considerations." *American Economic Review* 61: 112-23.

_____. 1973. "Markets and Hierarchies: Some Elementary Considerations." *American Economic Review* 63" 316-25.

_____. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York, NY: Free Press.

_____. 1979. "Transaction Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics* 22: 233-61.

_____. 1985. *The Economic Institutions of Capitalism*. New York, NY: Free Press.

_____. 1991. "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives." *Administrative Science Quarterly* 36: 269-96.

_____. 1996. *The Mechanisms of Governance*. Oxford: Oxford University Press.

_____. 2000. "The New Institutional Economics: Taking Stock, Looking Ahead." *Journal of Economic Literature* 38: 595-613.

_____. 2002. "The Theory of the Firm as Governance Structure: From Choice to Contract." *Journal of Economic Perspectives* 16: 171-95.

Winter, Sidney. 1988. "On Coase, Competence, and the Corporation." *Journal of Law, Economics, and Organization* 4: xx-yy.

_____. 2001. "Toward an Evolutionary Theory of Production." Unpublished manuscript, Wharton School, University of Pennsylvania.