Transaction Cost Economics*

Steven Tadelis and Oliver Williamson University of California, Berkeley November 14, 2010

Abstract

. . .

JEL classifications

^{*}We thank Bob Gibbons, Jon Levin and John Roberts for helpful comments on previous drafts. Financial support from the National Science Foundation grant number SES-0239844 is gratefully acknowledged.

1 Introduction

Throughout the past four decades the study of the governance of economic organization has evolved into a lively and diverse field of research. This chapter describes the fundamental ideas of Transaction Cost Economics (TCE), which emerged in the 1970's to offer a methodology through which to analyze how the governance of economic organization affects economic value. Our view, and the general outlook of TCE, is that, although this is an interdisciplinary project, organization matters for economists if and as organization is made susceptible to analysis by the application of economic reasoning.

As such, TCE is part of a broader effort to study the economics of organization, which includes agency/mechanism-design theory, team theory, property rights theory, and resource-based/competency theories. Many of these are explored in this handbook and, as appropriate, we make references to the similarities and difference between them and TCE. Lest we be misunderstood, we do not regard TCE as an all-purpose theory of firm and market organization. Instead, we subscribe to pluralism – on the understanding that all candidate theories spell out the phenomena to which they apply, and advance refutable hypotheses that lend themselves to empirical testing. Subject to the latter proviso, we are certain that a richer, deeper, better understanding of complex economic organization is well-served by the spirit of pluralism.

We begin in Section 2 by sketching early contributions upon which TCE builds and describe how TCE differs in three crucial respects from 1960's orthodoxy: moving from the *lens of choice* to the *lens of contract*; taking *adaptation* to be the main problem of economic organization; and taking *transaction cost*

economizing to be the main case for deriving refutable implications.

In Section 3 we lay out the working parts by which TCE was transformed into an operational methodology by taking the "make-or-buy" decision to be the focal transaction. We thereafter introduce the key idea that transactions, which differ in their attributes, will be best executed using different governance structures. Section 4 presents a simple formal model of TCE that parallels the less formal arguments of Section 3, wherein the choice of governance implements different contractual incentives. Extensions and applications — to include hybrid contracting, reality testing, variations on a theme, empirical TCE, and ramifications for public policy — are sketched in Section 5. Section 6 relates TCE to other economic theories of organization. Section 7 concludes.

2 Background and Essential Ideas

As Williamson (2010) explains, TCE traces its origins to a series of developments between 1930 and 1970 in economics (Commons, 1932; Coase, 1937, 1960; Hayek, 1945; Simon, 1951; Arrow, 1969), organization theory (Barnard, 1938; Simon, 1947; Selznick, 1949), contract law (Llewellyn, 1931; Summers, 1969), and business history (Chandler, 1962). Because this chapter is mainly directed to an economics audience and since organization theory, contract law, and business history are addressed elsewhere, we focus in this section, as well as this chapter, mainly on the economic arguments.

¹On organization theory, contract law, and business history, see Williamson (1993, 2005b, and 1985, chap. 11, respectively).

2.1 Transactions and Transaction Costs

Despite being largely ignored in textbook micro-economic theory, many transactions require parties to engage in a relationship over which ongoing interaction is needed to complete the transaction. John R. Commons recognized that beyond simple market exchange (e.g., exchanging nuts for berries on the edge of the forest or buying a can of coke at a vending machine), the continuity of an exchange relationship was often important. As he perceived the fundamental problem of economic organization, "the ultimate unit of activity ... must contain in itself the three principles of conflict, mutuality, and order. This unit is a transaction." (Commons, 1932, p.4) As we demonstrate in section 3, TCE follows Commons in that governance is chosen in a cost effective degree to infuse order, thereby to mitigate conflict and realize mutual gain. Furthermore, the transaction is made the basic unit of analysis.²

Ronald Coase, in his classic 1937 paper on "The Nature of the Firm," was the first to introduce the concept of transaction costs to the study of firm and market organization.³ Coase revealed a serious lapse in the accepted textbook theory of firm and market organization. Upon viewing firm and market as "alternative methods of coordinating production" (1937, p. 388), Coase observed that the decision to use one mode rather than the other should not be taken as given

²Whereas Commons viewed organization and the continuity of contractual relations as important, the economic paradigm of resource allocation, which dominated the field all through the 1970s, made negligible provision for either. Instead, It focused on prices and output, supply and demand. The institutional economics of the 1930s was mainly relegated to the history of thought because it failed to advance a positive research agenda that was replete with predictions and empirical testing.

³Though Coase did not use the term "transaction costs" in his 1937 article, he did suggest that employment relations were especially susceptible to transaction costs.

(as was the prevailing practice) but should be derived. Accordingly, Coase advised economists that they needed "... to bridge what appears to be a gap in [standard] economic theory between the assumption (made for some purposes) that resources are allocated by means of the price mechanism and the assumption (made for other purposes) that allocation is dependent on the entrepreneur-coordinator. We have to explain the basis on which, in practice, this choice between alternatives is effected." (1937, p. 389) Still, Coase's challenge failed to take hold over the next 20 years, during which period the standard assumption that transaction costs could be assumed to be zero went unchallenged. Two important articles in the 1960s would upset this state of affairs. Upon pushing the logic of zero transaction costs to completion, the unforeseen and disconcerting implications of this standard assumption were displayed for all to see.

Coase's 1960 article on "The Problem of Social Cost" reformulated the externality problem in contractual terms and pushed the logic of zero transaction cost reasoning to completion, with an astonishing result: "Pigou's conclusion (and that of most economists of that era) that some kind of government action (usually the imposition of taxes) was required to restrain those whose actions had harmful effects on others (often termed negative externalities)" was incorrect (Coase, 1992, p. 717). With no transaction costs parties will costlessly bargain to an efficient result whichever way property rights are assigned at the outset. Taken at face value, externalities and frictions of other kinds would vanish. But the real message was this: hereafter "study the world of positive transaction costs" (Coase, 1992, p. 717). Kenneth Arrow's 1969 examination of "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-market Allocation" likewise revealed a need to make a place for positive transaction costs, both with

respect to market failures and in conjunction with intermediate product market contracting: "the existence of vertical integration may suggest that the costs of operating competitive markets are not zero, as is usually assumed in our theoretical analysis" (1969, p. 48).

Accepting positive transaction costs, however, introduced three problems. First, upon opening the "black box" of firm and market organization and looking inside, the black box turned out to be Pandora's Box: positive transaction costs were perceived to be everywhere. Since some form of transaction cost could be invoked to explain any condition whatsoever after the fact, the appeal to transaction costs acquired a "well deserved bad name" (Fischer, 1977, p. 322). Second, transaction costs take on comparative institutional significance only as they can be shown to differ among modes of governance (say, as between markets and hierarchies). Third, transaction costs that pass the test of comparative contractual significance need to be embedded in a conceptual framework from which predictions can be derived and empirically tested. The unmet need was to focus attention on key features and provide operational content for the intriguing concept of positive transaction costs.

2.2 Governance and the Lens of Contract

James Buchanan distinguished between lens of choice and lens of contract approaches to economic organization and argued that economics as a discipline went "wrong" in its preoccupation with the science of choice and the optimization apparatus associated therewith (1975, p. 225). If "mutuality of advantage from voluntary exchange is ... the most fundamental of all understandings in economics" (Buchanan, 2001, p. 29), then the lens of contract approach is an

under-used perspective. Thus, in order to analyze the inner working of institutions such as markets and hierarchies, the orthodox lens of choice (the resource allocation paradigm, with emphasis on prices and output, supply and demand) would need to give way to the evolving lens of contract. This position was implicit in Commons (1932) and Coase (1937) and was explicit in Coase (1960) and Arrow (1969). Upon examining governance and organization through the lens of contract, the firm was no longer a black box for transforming inputs into outputs according to the laws of technology but was interpreted instead as an alternative mode of contracting.

Indeed, this shift to the lens of contract was adopted not only by TCE but more generally – in mechanism design, agency theory, information economics, and formal property rights theory. But there were also differences (described in more detail in Section 6). Whereas all of the latter located the analytical action on ex ante incentive alignment, TCE concentrated the main action on the ex-post governance of contractual relations. This emphasis on ex post governance is congruent with TCE's taking adaptation to be the main problem of economic organization.

2.3 Adaptation as the Central Problem

Friedrich Hayek (1945) suggested that "economic problems arise always and only in consequence of change" (1945, p. 523), whereupon "rapid adaptation to changes in the particular circumstances of time and place" was taken to be the main problem of economic organization (1945, p. 524). Interestingly, Chester Barnard (1938), who was an organization theorist rather than an economist, likewise regarded adaptation as the central problem of organization. But while Barnard and Hayek were in agreement on the importance of adaptation, each had reference to different types

of adaptation.

Hayek focused on the adaptations of autonomous economic actors who adjusted spontaneously to changes in the market, mainly as signaled by changes in relative prices. An illustrative recent example is the increase in the demand for fossil fuels that caused gasoline prices to rise sharply in 2008. As a result, automobile firms adapted their strategies to invest in alternative energy vehicles and create new products, which in turn will affect the demand for fuel. Prices will shift to equilibrate the new choices, and consumers and producers will adapt their consumption and production behavior as part of the equilibrium. The marvel of the market thus resides in "how little the individual participants need to know to be able to take the right action" (1945, pp. 526-527).

By contrast, Barnard featured coordinated adaptation among economic actors working through administration. In his view, the marvel of hierarchy is that coordinated adaptation is accomplished not spontaneously but in a "conscious, deliberate, purposeful" way through administration (1938, p. 4). As an illustration, Boeing experienced major coordination problems in the last quarter of 2007 because its outside suppliers were preoccupied and sometimes overwhelmed with their own problems to the neglect of systems considerations. Serious delays in the production of their most important future aircraft, the 787 Dreamliner, resulted. Boeing, in effect, had made major bets on the efficacy of outsourcing on a scale that they had never previously attempted, but more attention to the needs for coordinated adaptation to which Barnard referred was evidently needed. Boeing acquired one of their most important suppliers in the summer of 2009 precisely for this reason.

TCE concurs that adaptation is the central problem of economic orga-

nization and makes provision for adaptations of both autonomous and coordinated kinds. The "marvel of the market" (Hayek) and the "marvel of hierarchy" (Barnard) are now therefore joined. The upshot is that the problem of economic organization is properly posed not in terms of the old ideological divide of markets or hierarchies but rather as the combined use of markets and hierarchies.

2.4 Efficiency

TCE primarily holds that adaptations of both kinds are undertaken in the service of efficiency. Specifically, the choice among alternative modes of governance mainly has the purpose of economizing on transaction costs.

However, although TCE claims that nonstandard and unfamiliar contracting practices mainly operate in the service of efficiency, other purposes, of which market dominance (monopoly) is one, are sometimes responsible for puzzling practices. In fact, in the pre-TCE era, monopoly purpose was believed to be mainly responsible for nonstandard and unfamiliar contracting practices and modes of organization. As discussed in Section 5.5, this preoccupation with monopoly was wrong-headed and gave rise to convoluted antitrust enforcement, which in the 1960s sometimes spun out of control. (This was one of the factors that prompted a re-examination of vertical integration in comparative transaction cost economizing terms (Williamson, 1971).)

3 Making TCE Operational: Vertical Integration

As developed elsewhere (Williamson, 1971, 1975, 1985), and as described in this section, TCE was made operational in three steps. First, it took the transaction to be the basic unit of analysis and named the key attributes across which transactions differ. Second, it described the properties of alternative modes of governance. Last, the analysis was completed by applying the "discriminating alignment" hypothesis: different kinds of transactions are more efficiently governed by different modes of governance.

Conceivably, these ideas could be implemented in the abstract. But rather than develop a "general theory" from the outset, TCE took shape by examining the particulars. The choice of vertical integration (more generally, the intermediate product market transaction – the "make-or-buy" decision) as the focal transaction was not only the obvious candidate by reason of the antitrust concerns referred to above, but it would turn out to be fortuitous: it led quickly and easily into an interpretation of many other phenomena as variations on a theme.

The other obvious candidate to serve as a focal transaction was the employment relation (as in Coase (1937) and Simon (1951)). Coase explained his choice of the employment relation as follows: "A firm is [more] likely to emerge... in the case of services – labor – than it is in the buying of commodities" since, for the latter, "the main items can be stated in advance and the details which will be decided will be of minor significance" (1937, p.392). But while that is true for some intermediate product market transactions, it is by no means true for all. Indeed, it is precisely because of the variations among intermediate product market trans-

actions that firms reach different decisions as to make-or-buy. Establishing where these differences reside was crucial to the transaction cost economics project – both with respect to vertical integration and, more generally, to the interpretation of other transactions (labor included) as variations on a theme.⁴

Once a focal transaction has been decided, the three key operational moves are to (1) name the attributes of the unit of analysis, (2) do the same for modes of governance, and (3) advance the efficient alignment hypothesis.

3.1 Unit of Analysis: Classifying Transaction Attributes

The obvious place to start is with the ideal transaction of simple market exchange where identity does not matter and the terms of trade are completely specified. Competition and market mediated exchange can be presumed to prevail in such circumstances. A simple example would be the market for standard thread screws – these well defined simple commodities are readily and competitively supplied through anonymous spot market transactions.

Whereas simple market exchange works well for transactions where identity does not matter, we consider the attributes of transactions where identity does matter and for which added governance might be warranted. Longer term contracts through which the parties value continuity come under scrutiny in this way.

⁴Coase appealed to labor law to support the primacy of labor: because "the fact of direction is the essence of the legal concept of 'employer and employee' just as it was in the economic concept which was developed above ... [we can] conclude that the definition we have given is one which approximates closely to the firm as it is considered in the real world" (1937, p. 404).

3.1.1 Asset specificity and the Fundamental Transformation

One of the early innovations of TCE was to give prominence to the relatively neglected condition of asset specificity as a crucial defining attribute of transactions. Asset specificity describes the condition where the identity of the parties matters for the continued relationship, and it can take a variety of forms – physical, human, site specific, dedicated, brand name capital, and episodic (sometimes described as temporal specificity) – the optimal response to which varies somewhat but involves greater reliance on "administration." Whatever the form, these assets cannot be redeployed to alternative uses or users without loss of productive value (Williamson, 1971, 1975, 1976, 1985; Klein, Crawford, and Alchian, 1978). Asset specificity can sometimes arise spontaneously, without conscious and costly investments, as with knowledge and skills that are incidentally acquired by the parties while working together.

By transaction specific assets we mean ones for which a bilateral dependency relationship develops between the parties. Even though a large number of qualified suppliers could compete for the job before the contract is awarded, the relationship is effectively transformed during contract implementation into a bilateral supply relation thereafter. Asset specificity thus gives rise to a Fundamental Transformation in the contractual relation as bilateral dependency builds up.⁵ Identity thereafter matters.

⁵Note that because asset specificity is a design variable, the good or service to be delivered could be redesigned by reducing asset specific features, albeit at a sacrifice in performance of the good or service in question (Riordan and Williamson, 1995).

3.1.2 Uncertainty, Complexity and Incomplete Contracts

Asset specificity, in any of its forms, does not by itself pose contractual hazards that require added governance. In principle, a long-term contract can offer the parties adequate compensation for efficient investment in specific assets, and spell out how future contingent decisions that rely on these assets will be made and compensated for.

Contractual hazards will, however, arise if and as long-term contracts are incomplete and are subject to (1) disturbances (or contingencies) during the contract implementation phase for which it is prohibitively costly to prescribe appropriate responses in advance, requiring ex post adaptation of the original plan, (2) defection from the spirit of cooperation to insist on the letter of the contract can be projected for some extreme disturbances (where the stakes are great), and (3) the courts cannot be relied upon to fill gaps and settle disputes in a timely, knowledgeable, and efficient fashion.

Accordingly, in addition to a condition of bilateral dependency (a consequence of asset specificity), the contractual hazards for which added governance supports are needed turn on (1) uncertainty and contractual incompleteness,⁶ (2) strategic defection (when the stakes are large), and (3) the serious limits of court ordering – where this last presumes that self imposed modes of governance (private ordering) can sometimes be devised and implemented with expected net gains.

Larger degrees of a transaction's complexity and uncertainty, and the inability to describe upfront the appropriate responses to such contingencies, lead to

⁶Incompleteness becomes more severe as the number of features of transactions (precision, linkages, compatibility) across which adaptations are needed increases and as the number of consequential disturbances that impinge upon these features increases. These features can be thought of as a measure of complexity (see Section 4.1.2).

more consequential contractual incompleteness (gaps, errors, omissions, and ambiguities arise). Even, moreover, if it were feasible to contract comprehensively, the prohibitive cost of complete contracts would discourage such efforts (e.g., Dye (1985), Bajari and Tadelis (2001)). Being impossible or prohibitively costly to contract comprehensively, many contingencies are dealt with in an adaptive, sequential way (Williamson, 1971, 1975, 1985). This, however, is subject to problems when a bilaterally dependent buyer and supplier lack the incentives to cooperate.

On the assumption that contractual incompleteness, strategic propensities, and the limits of court ordering are all consequential, asset specificity and disturbances for which unprogrammed adaptations are needed are the two crucial attributes of transactions for understanding the governance of contractual relations.⁷

3.2 Governance Structures

As argued above, the central problem of adaptation becomes more severe as assets are more specific and contracts are more incomplete. A promising solution is to introduce contractual governance safeguards that mitigate these hazards. This poses the challenge of describing well defined alternative modes of governance and explaining the strengths and weaknesses that are associated with each. Following our discussion in Section 2, we focus our analysis on the two conventional polar

⁷A related factor, which we pass over here because of its ambiguous implications, is the frequency with which transactions recur. On one hand, if a transaction seldom recurs, it may not be cost effective to develop a specialized internal structure. If instead it recurs frequently, then recovering the costs of creating a specialized management infrastructure is possible. On the other hand, recurrent contracting implies that future business is at stake, for which a good reputation figures in. Market contracting, if supported by good reputation effects, thus becomes part of the comparative contractual calculus.

modes of governance, markets ("buy") and hierarchies ("make").

TCE maintains that each polar mode of governance – market and hierarchy – is described as an internally consistent system of attributes with distinctive strengths and weaknesses. Williamson (1985) — identified two main attributes by which each mode of governance is described. (1) incentive intensity (which is measured by the extent to which a technologically separable stage of economic activity appropriates its net profits), (2) administrative authority and control (which has a bearing on autonomy in both operating and investment respects as well as on procedural controls (routines, accounting procedures)). More implicit in the original treatment (Williamson, 1985) was a third attribute: the role played by contract law regime. This third integral part of the picture receives a more serious treatment in Williamson (1991) and is summarized in Williamson (2010). Our focus in this chapter, especially in the formal model of section 4, focuses attention on the first two attributes: incentive intensity and administrative control.⁸

We will take incentives intensity and administrative control to be the key attributes and will assume for now that each can take on either of two values: strong or weak. Williamson (1985, chapters 4 and 6) argues that (by reason of the impossibility of implementing replication with selective intervention – see section 3.4 below) the viable combinations are (1) strong incentive intensity and weak administrative controls at the interface and (2) weak incentive intensity and strong

⁸The important consideration of supporting legal regimes is set aside due to the challenge of modeling this aspect of governance. By and large the formal literature—be it mechanism design or the property rights approach—assumes that whatever is contractually verifiable will be enforced by some unmodeled third party. In this regard one can view this chapter as a bit schizophrenic: we recognize here that contract law regime can and will vary along the choice of governance, yet in section 4 we follow the standard modeling approach and focus only on incentive intensity and administrative control.

administrative controls at the interface. The first of these corresponds to the (spot) market, which entails outsourcing for the good or service in question. Because each stage is independently owned and operated, each appropriates its own net receipts (high powered incentives) and adapts promptly when confronted with disturbances without heed for the other (weak interface controls). The second corresponds to hierarchy, where the two stages are commonly owned and (under low powered incentives) will respond without resistance to interface instructions to adapt coordinately (strong administrative controls). Plainly, whether to outsource to the market (buy) or produce to one's own needs (make) will vary with the perceived adaptive needs of the transactions in question.

Two other configurations are possible. Strong incentives at each stage and strong administrative control at the interface between stages spells conflict: if each stage accrues its net receipts then incentives to follow the implied direction of strong administrative control at the interface towards cooperative adaptation will be resisted.

Finally, weak incentives at each stage and weak administrative control at the interface leads to a compliant supplier relation akin to cost-plus outside procurement where, within some bounds, the supplier does what the buyer asks because it is assured of reimbursement, although some auditing by the buyer will usually attend such relations.⁹ Table 1 summarizes the four implied governance modes.¹⁰

⁹For a formal treatment of relative benefit of cost-plus contracts over fixed price contracts with respect to adaptation see Bajari and Tadelis (2001).

¹⁰To briefly relate legal rules to the choices of incentives and administrative control, contract law regime must be supportive of the attribute that is strong (incentive intensity or administrative control). For strong incentives to be effective, parties should have control over their processes and the letter of the contract should be expected o be enforced (strong legal rules). In contrast,

		Incentive	Intensity
		Strong	Weak
Administrative Control	Weak	Market (buy)	Cost-Plus
			Contracts
			(less common)
	Strong	Recipe for	Hierarchy (make)
		Conflict	
		(empty)	

Table 1: Alternative Modes of Governance

As suggested above, market mediated exchange takes place between independently owned and operated stages whereas hierarchy mediated exchange is accomplished by unified ownership and coordinated interface mediation. These operating differences are truly consequential and warrant elaboration.

TCE takes adaptation to be the central problem of economic organization of which two types are distinguished: Hayek's autonomous adaptation (the market) and Barnard's coordinated adaptation (hierarchy). With high-powered incentives and the absence of administrative involvement, the parties to market mediated exchange will negotiate the initial contract in a hard-headed way. Each will appropriate the stream of net receipts that accrues to it from autonomous adaptation during the execution of the contract, and either party can appeal disputes to the courts, which in turn will apply the appropriate legal rules to award money damages. This is shown in the upper panel of Figure 1 as market mediated exchange.

forbearance law contract regime (private ordering with weak legal rules) is supportive of strong administrative control since interference by courts would undermine the value of having control by fiat. For more on this see Williamson (1991).

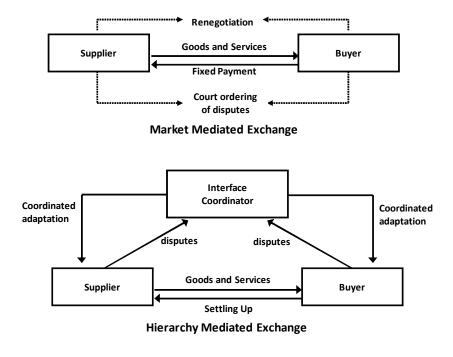


Figure 1: Interface Mediation: Market Versus Hierarchy

By contrast, coordinated adaptation under hierarchy, as shown in the lower panel of Figure 1, is promoted by unified ownership of the two stages coupled with the creation of a new actor, the interface coordinator, to which each stage reports and receives administrative direction and control. Consequential disturbances that would give rise to poor coordination are dealt with by the interface coordinator, who has ultimate responsibility for coordinated responses; and internal disputes between stages are likewise settled by the interface coordinator with reference to mutual gain (private ordering). ¹¹

¹¹To tie this briefly to legal rules and regimes, the efficacy of internal dispute resolution and of coordinated adaptations, as decided by the interface coordinator, would be severely compromised if disgruntled managers could go over the coordinators' heads and appeal to the courts.

3.3 Efficient Alignment

Implicit in the preceding discussion is the "discriminating alignment hypothesis" to which TCE owes much of its predictive content. Namely, transactions, which differ in their attributes, are aligned with governance structures, which differ in their adaptive capacities, so as to minimize transaction costs. Note that (1) markets enjoy the advantage of autonomous adaptation for transactions that are supported by generic assets, (2) the importance of coordinated adaptation increases as the supporting assets become more specific, and (3) the bureaucratic costs of hierarchy, among them the loss of incentive intensity, are a deterrent to integration except as coordinated adaptation benefits are more than offsetting.

If contracts are incomplete and some consequential disturbances are expected, efficient alignment is thus accomplished by assigning transactions for which asset specificity is low to markets and transactions for which asset specificity is great are assigned to hierarchy. Using σ as an index of asset specificity, the efficient alignment argument is displayed in Figure 2.¹² Thus, if one can measure the attributes of transactions, namely asset specificity and the likelihood of consequential disturbances, efficient alignment offers a recipe for governance that is empirically testable.

3.4 The Impossibility of Combining Replication with Selective Intervention

Sections 3.1-3 outline the main logic of transaction cost economics as applied to the make-or-buy problem. But there has been a long standing puzzle posed by

¹²In this binary setup, transactions for which asset specificity is in the neighborhood of ■* can go either way, since it does not matter very much; but see the discussion of hybrids in Section 5.1.

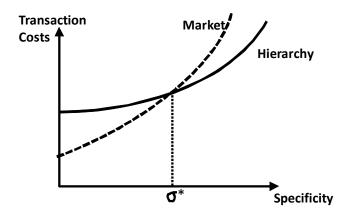


Figure 2: Efficient Governance

Frank Knight (1921, 1931) and Ronald Coase (1937): Why can't a large firm do everything that a collection of smaller firms can do and more? Tracy Lewis (1983, p. 1092) argues that because an established firm can always "use the input exactly as the new entrant would have used it . . . [and can furthermore] improve on this by coordinating production from his new and existing inputs" then the large firm will indeed always realize greater value. Transaction cost economics takes exception with this argument by examining the efficacy of the two mechanisms on which Lewis's formulation implicitly relies: replication and selective intervention.

Imagine, without loss, that the buyer stage acquires the supplier stage with the understanding that (1) the supplier will operate in the same autonomous way post-acquisition as in the pre-acquisition status (by replication) except as (2) the buyer intervenes selectively when expected net gains can be achieved through coordinated adaptations. The combined firm can then never do worse (by replication) and will sometimes do better (by selective intervention). Therefore, more integration is always better than less, and repeating this logic implies that everything will be organized in one large firm.

To see where this logic breaks down one must examine closely four implicit conditions for implementing replication and selective intervention: (1) the buyer (owner) promises the acquired supplier its net receipts in all state realizations – thereby preserving strong incentives; (2) the supplier promises to use the supply stage assets that the buyer now owns with "due care"; (3) the buyer promises to exercise authority (fiat) only when expected net gains can be ascribed to selective intervention; and (4) the buyer promises to reveal and divide the net gains from selective intervention as stipulated in the original agreement. The problem is that none of these promises is self-enforcing. To the contrary, in the absence of perfect information (to include a costless arbiter), each condition will be compromised. A detailed account of why each of the four conditions above fails is reported elsewhere (Williamson, 1985, Chap. 6).¹³

The upshot is that the action resides in the details and that these need to be uncovered and examined in a comparative way – which reveals that markets and hierarchies differ in discrete structural ways, each with its strengths and weaknesses.

¹³Contributing factors include (1) the owner (buyer) controls the accounting system and, within limits, can declare depreciation, transfer prices, and benefits so as to shift net receipts to its advantage, (2) failures of due care become known only with delay and are difficult to prove, (3) the buyer can also falsely declare state realizations to favor its own stream of net receipts, and (4) in consideration of the foregoing, the division of benefits under selective intervention can be compromised. Also, (5) the political game is now played in a larger firm that is more susceptible to bureaucratic ploys and political positioning than in smaller firms. Suffice it to observe that the breakdowns to which we refer are not beyond the grasp of intelligent businessmen and their lawyers, who recognize the tradeoffs and factor them into the integration calculus.

4 A Simple Governance Model

Building upon Tadelis (2002), we construct a model of contractual choice that parallels the arguments in Section 3. Despite its extreme reduced form, the model relies on micro-foundations developed by Grossman and Hart (1986), Holmstrom and Milgrom (1991), and Bajari and Tadelis (2001).¹⁴ The model contributes by exposing complementarities between cost incentives and governance that helps clarify the underpinnings of the TCE framework outlined in section 3.

The model parallels Section 3 in that (1) exchange takes place between successive (technologically separable) stages of production; (2) spot markets aside, all contracts are incomplete to varying degrees; (3) the critical attributes of transactions are asset specificity and contractual incompleteness (disturbances), where the former is responsible for bilateral dependency and the latter creates the adaptive need; (4) if the parties are independent and if a disturbance occurs for which the contract is not adequate, adaptation is accomplished by renegotiation and/or court ordering; and (5) efficiency is served by aligning transaction with governance structures in a transaction cost economizing way.

To simplify, however, there are two differences between the formal model and Section 3 that are worth noting, and more discussion will follow. First, adaptation costs are incurred only by the buyer. (The efficient alignment result does not depend on this asymmetry.) Second, asset specificity is treated as a probability of finding an alternative seller without incurring adaptation costs, rather than as an actual loss in surplus if the seller is replaced. This is done for simplicity purposes.

¹⁴An early step in formalizing transaction cost economics can be found in Riordan and Williamson (1985).

4.1 Transactions and Governance

Consider a transaction where a buyer achieves value v > 0 if it procures a good (or service) from a supplier and successfully incorporates it into its own output. The transaction is characterized by the degrees of asset specificity and contractual incompleteness.

4.1.1 Asset Specificity

Recall from section 3.2 that asset specificity can manifest itself in several ways: physical, human, site, dedicated assets and brand name capital. Asset specificity can arise from both purposeful investments and spontaneously, where the latter take the form of knowledge and skills that are incidentally acquired by the parties while working together. Whatever the form, these assets cannot be redeployed to alternative uses and users without some loss of productive value.

Asset specificity is modeled by $\sigma \in [0,1]$, where higher values of σ represent higher degrees of asset specificity. It will be technically convenient to interpret σ as the probability that the supplier cannot be replaced by a competitor when disturbances occur, and adaptation costs will be incurred. With probability $(1-\sigma)$, however, there exists an alternative supplier who will compete to perform any adaptations that are needed and adaptation costs will be avoided.

Though we believe that more asset specificity does imply that more value is lost when an alternative supplier is used, from an ex ante perspective the probabilistic nature of our measure of asset specificity is equivalent: when σ is higher, the expected loss from having to switch suppliers will be larger, making the fundamental transformation more severe.

4.1.2 Complexity and Incomplete Contracts

Contractual completeness is modeled by $\tau \in [0, 1]$, interpreted as the probability that the contract will adequately cover ex post needs. With probability $(1 - \tau)$ a significant disturbance (or contingency) occurs, at which point the contract's ex ante design will fail to achieve the value v. In this event, ex post adaptation at some extra cost will be needed to achieve v.¹⁵ Bajari and Tadelis (2001) consider a model where it is possible to invest more or less resources in design ex ante, making design (i.e., contractual completeness) an endogenous variable that responds to project complexity. They show that more complex projects have endogenously chosen contracts that are more incomplete. Hence, we treat $(1 - \tau)$ as exogenous and interpret it as the *contractual incompleteness* of the transaction, a characteristic discussed in section 3.2.

4.1.3 Markets and Hierarchies

Two modes of governance are considered: market and hierarchy. Recall from Section 3.2 that TCE identifies markets with two features: first, both buyer and supplier face high-powered cost incentives in that they each receive their own net receipts, and second, each maintains administrative control over their own production process and adaptations must be renegotiated. In contrast, TCE identifies hierarchy with the opposite features: first, the parties face low-powered cost incentives in that their costs are reimbursed, and second, the parties relinquishes administrative control over processes to an "interface-coordinator". We proceed to formally define market versus hierarchy using only one of these attributes: the

 $^{^{15}}$ For simplicity we keep v constant so that any loss from adaptation is captured by adaptation costs, to be defined precisely later.

assignment of administrative control over production and adaptation processes. The strength of incentives will be separately derived, and the model will endogenously result in outcomes that match the observations in section 3.¹⁶

Define market governance (M) to be the choice in which the parties retains autonomy over their own production process decisions, and the supplier is expected to deliver a product that meets the contractual specifications. Any adaptation to adjust the ex ante design due to disturbances needs to be renegotiated by the the autonomous parties. For simplicity, we assume that adaptation will be required only for the supplier's production phase of the project. Our asymmetric treatment of disturbances simplifies the analysis and offers a first step in formalizing the ideas described in section 3. Naturally, the buyer may have to adapt the process under its control to accommodate some disturbances as well. Symmetric disturbances on both sides of the transaction will result in a more cumbersome analysis, obscuring the main insights.

Define hierarchy (H) to be the choice in which the parties relinquish administrative control to a third party, the interface coordinator. This means that routine tasks are followed as planned, but when disturbances arise then decisions are made by the interface-coordinator who possesses unified ownership and control over production and adaptation for both stages of buyer and supplier.

It is important to note that our notion of hierarchy as unified control and coordinated adaptation differs from a "directional" integration argument of whether the buyer integrates the supplier into its business and becomes the interface-

¹⁶The model ignores conflict resolution law and follows the common approach in contract theory that the courts will enforce whatever the contract stipulates. This is also the approach taken by Grossman and Hart (1986). It is important to note, however, that current business law, which is very much commensurate with the brief description in Section 3.3, would further strengthen the conclusions of our theoretical modelling exercise.

coordinator, or the reverse, which is a novel and central feature of the Property Rights Theory (PRT) that is presented in chapter X of this handbook. In contrast, our notion of integration assigns responsibility for implementing routines to the managers at each stage, whereas disturbances for which coordinated adaptations are required are done at the direction of the interface coordinator (see Figure 1). In a market transaction, however, each party controls their own processes, and consent between buyer and supplier are needed for any adaptation. TCE identifies this interface-coordinator as often being a third party whose incentives are aligned with total profit maximization. That is, instead of a preexisting buyer and supplier, the transaction is a de novo investment whose governance needs to be determined. Efficiency considerations will determine whether the transaction is integrated (controlled by a interface-coordinator) or if it is not integrated (controlled by the contract and mutually agreed upon adaptations). PRT, in contrast, identifies integration with the situation in which one of the two parties becomes the owner of all productive assets, and controls the decisions related to their use. The predictions of PRT are as much about which of the two parties maintains control as about when unified ownership is called for.

Aside from the allocation of administrative control, a compensation scheme that the buyer (or interface-coordinator) uses to compensate the supplier must also be chosen as part of the inter-firm contract (or intra-firm compensation scheme). This will influence the supplier's incentives to reduce costs. Denote the supplier's production costs by c, which includes material and other expenses (lost opportunities, possibly the wages of laborers under its direction, etc). We restrict attention to linear compensation schemes that have a fixed component, F, and a share of production costs, $(1-z) \in [0,1]$. A supplier that incurs cost c is paid F + (1-z)c,

where $z \in [0, 1]$ is the share of production costs that are borne by the supplier, often referred to as the "strength" of cost-incentives that the supplier faces. For example, if z = 1 and F > 0 then the supplier receives a "fixed-price" payment of F and bears all of the production costs, which seems to be the norm for almost all market transactions. In this case the supplier has very strong incentives to find ways to reduce production costs. In contrast, if z = 0 and F > 0 then a "cost-plus" contract is in place where the supplier receives some fixed compensation F and bears none of the production costs c, which seems to be the case for most hierarchical structures (of course, in this case the chosen level of F will be different from the case in which z = 1). With a cost-plus contract the supplier has no gain from engaging in activities that reduce production costs. Both F and z will be endogenously derived in our analysis.

4.2 Production Costs

Assume that the supplier's production costs are given by the function c(e, G), where $e \geq 0$ is the effort intensity that the supplier puts into the project, and $G \in \{M, H\}$ denotes the mode of governance. More effort intensity results in lower production costs. This can be the extra time and attention that the supplier puts into directing production of other employees, into choosing production alternatives that reduce costs, etc. The (opportunity) cost of effort to the supplier is given by a convex increasing function y(e). We impose the standard agency assumption that contracting on effort intensity is impossible.

Recall that the choice of governance determines whether decisions for disturbances will be made autonomously by the buyer and supplier, or by an interface coordinator. Arguably, *decisions* on how to produce or adapt the production

process can be thought of as an *input* into the production process. If the supplier has complete control of its production process then the ability to control all the inputs, including effort *and* decisions (as is the definition of markets), should make the supplier more effective in cutting costs. If, however, decisions are in the hands of the interface coordinator (as is the definition of hierarchy), then the supplier's lack of autonomy will make its effort less effective. Formally, we assume:¹⁷

A1: Supplier effort is more effective in market governance: $\frac{\partial c(e,M)}{\partial e} < \frac{\partial c(e,H)}{\partial e} \leq 0$.

In addition to the intuitive appeal described above, A1 is a consequence of previously established micro-foundations.¹⁸ Note that A1 is an assumption about a marginal reaction, namely, the effect of the allocation of administrative control on the slope of the cost function with respect to effort. Arguably, the level of production costs ought to be lower when the supplier retains control, i.e., $c(e, M) \leq c(e, H)$.¹⁹ If we assume that when the supplier exerts no effort then the allocation

¹⁷Expanding the model to have buyer effort and a role for buyer incentives is an easy extension, with each stage's effort intensity being more effective for that stage when the stage has control over all aspects of production. Our asymmetric simplification relates this to the supplier stage alone.

¹⁸First, Grossman and Hart (1986) model ex post decisions that need to be made after a state of nature is realized, yet contingent contracts cannot be written ex ante. The state of nature is very much like our ex post disturbances that are not contracted for. If the decisions themselves affect the way in which effort reduces production costs, then when the supplier controls both decisions and effort, he is better able to reduce his costs. Ownership of productive assets—as defined by PRT—gives the supplier control over the allocation of productive assets ex post, and this makes his cost-reducing effort more effective. Indeed, this assumption is our parallel of Assumption 6 in Hart and Moore (1990). Second, A1 is also consistent with the models of Holmstrom and Milgrom (1991, 1994). They show that the supplier's effort may be misdirected when he does not own the productive assets, where ownership is associated with having control over the use and sale of assets. If the productive assets are not owned by the supplier then he will not necessarily take the asset's long run value into account, imposing additional long-run costs on production. (This idea is discussed explicitly in Williamson, 1985, pp. 137-138.)

¹⁹Note that we are only referring to production costs when the contract adequately described

of administrative control does not matter (i.e., c(0, M) = c(0, H)) then A1 implies that c(e, M) < c(e, H) for all e > 0.20

4.3 Adaptation Costs

If a disturbance occurs then above and beyond the production costs c(e, G), costly adaptations must be made to obtain the value v. Adaptation costs can have at least two sources. The first will be imposed by having some previous activities wasted and redone, or having to modify initially planned production processes that fit the original design. These adaptation costs stem from contractual incompleteness and could have been spared if a complete contract and accurate design were in place. Yet as posited, providing such details was either impossible or, in expectations, not cost effective. The second source of adaptation costs will be a result of haggling, rent seeking and other renegotiation costs that parties expend in order to get a better deal. These activities are a source of dead-weight loss that does not affect equilibrium outcomes.²¹ We are agnostic as to which of these two sources of

the transaction. We will next turn to what happens when adaptation is required.

²⁰Notice that we make use of the method developed by Grossman and Hart (1986) to model the effect of administrative control. They modelled changes in the ownership of assets as changing the marginal effect of effort. Despite that the driving forces and economic insights that we derive are different than those in PRT, the PRT modelling method will prove useful in formalizing the original ideas developed by TCE.

²¹These haggling costs can be generated from a simple rent seeking game in the spirit of Tullock (1980) or Skaperdas (1992). As a simple example imagine that the surplus to be split is 10 and that without hiring lawyers, a court (or arbitrator) would direct the parties to split the surplus equally. Assume that if a party hires a lawyer to argue in court then he would get $\frac{3}{4}$ of the surplus if the other party has no lawyer, but symmetrically, would get $\frac{1}{2}$ of the surplus if the other party has a lawyer. If the lawyer costs are k < 2.5 then this game a prisoner's dilemma in which hiring a lawyer is a dominant strategy, the parties split the 10 equally but each pays k and is left with 5 - k. Hence, as far as the relationship is considered, rents were dissipated. This idea is explored in Ashenfelter and Bloom (1993) who show that parties to disputes indeed have an incentive to engage in legal representation, and this suggests that some excessive legal

adaptation costs are more or less severe, and we will aggregate them together as total adaptation costs denoted by k(z, G) > 0.

Adaptation costs k(z, G) are incurred if two events happen. First, a disturbance must happen, which occurs with probability $(1 - \tau)$, the incompleteness of the contract. Second, a new supplier from the competitive market cannot do the work, which occurs with probability σ , our measure of specificity. The rationale for this is simple: if the needed adaptation is not very specific then it would be more likely that other suppliers can compete for the adaptation that needs to be done, and no loss from haggling will occur.

As mentioned several times, we simplify and assume that adaptation costs are borne only by the buyer. We can easily have the supplier also bear some adaptation costs, which would add symmetry to the problem of adaptation without changing the qualitative results.²² We also assume that adaptation costs are worth incurring, so that when adaptation is needed the buyer's gross benefit is given by v-k(z,G) > 0. Hence, the expected gross benefit is given by $v-(1-\tau)\sigma k(z,G) > 0$, and expected adaptation costs are increasing in both contractual incompleteness, $(1-\tau)$, and asset specificity σ .²³

expenses may be expended.

²²For instance, if ex-ante competition pushes expected rents ex ante to zero, then any adaptation costs (or renegotiation income) that the supplier expects to spend (or gain) will be incorporated into his bid and will fall back on the buyer (see, e.g., Bajari, Houghton and Tadelis, 2009).

²³One can model the effects of adaptation as lowering the payoff v to some v' < v, above and beyond any additional adaptation costs of k, as in Tadelis (2002). This is redundant because the expected gross benefit (ignoring production costs) to the buyer would be $\tau v + (1-\tau)[(1-\sigma)v + \sigma(v'-k) = v - \sigma(v-v'+k)]$, and since v-v'>0 then this is just part of the adaptation costs as we define them above. Also, it is without loss that we assume that adaptation is always efficient. It is easy to incorporate a more involved setting where adaptation costs are random, and that they may sometimes be inefficient in that some projects are sometimes terminated. (See, e.g., Bajari and Tadelis, 2001.)

The choice of governance affects adaptation costs.²⁴ The bilateral dependence causes a conflict of interest between the buyer (or the interface-coordinator) and the supplier when adaptation is required. If the supplier has autonomy (market) then it is in a stronger position to hold-up the buyer and direct activities in its interest. Furthermore, if the supplier controls the adaptation process then it has more power to choose actions that direct rents towards it, which would impose added costs on the buyer and may even reduce total surplus. In a hierarchy, however, the interface coordinator has control over the adaptation process. This reduces the adaptation costs that the supplier can impose.²⁵ Summarizing:

A2: Adaptation costs are lower in hierarchy: k(z, M) > k(z, H) for all z.

The intuitive appeal of A2 is further supported by the micro-foundations of the hold-up problem. The more specific assets of any kind are concentrated under an independent supply stage, the worse is the fundamental transformation. When the supplier owns the dedicated assets then the *temporal lock-in* is stronger because the aggregate degree of bilateral dependency has increased.²⁶

We now argue that the choice of incentives will affect adaptation costs. Williamson proposed that "low powered incentives have well known adaptability advantages. That, after all, is what commends cost plus contracting." (1985 p.140). Indeed, if the supplier bears less of the production costs then it has less to gain from engaging

²⁴This general idea appears already in Williamson (1985) chapter 6, and is also realted to the work of Milgrom and Roberts (1988).

²⁵In a symmetric model where the buyer can impose adaptation costs on the supplier (or interface coordinator) then with autonomy, the buyer can impose adaptation costs to redirect rents towards himself. Adaptation costs imposed by both sides would be reduced under hierarchy.

²⁶Furthermore, consistent with the assumptions of PRT, giving the supplier ownership over assets (or simply control over adaptation decisions) will strengthen its bargaining position (outside option) and increase its renegotiations rents.

in negotiations with the buyer (or interface-coordinator) who is requesting some adaptation. Formally:

A3: Adaptation costs are lower when cost-incentives are weaker: $\frac{\partial k(z,G)}{\partial z} > 0$.

The micro-foundations for A3 are explored in Bajari and Tadelis (2001). In a buyer-supplier contracting model with bilateral dependency they show that if the supplier's incentives are stronger then the rents dissipated during adaptation are larger. Intuitively, when the supplier incurs more of the costs then the conflict of interest between buyer and supplier is intensified: the buyer wants a change that maximizes its net benefit, while the supplier will be motivated by reducing its costs. If the supplier bears less of the costs then the conflict is weaker and adaptation costs are lower. This insight should carry over to an integrated setting where an interface-coordinator directs adaptation: if the interface-coordinator mediates both stages of production and seeks to maximize combined surplus then both supplier and buyer will impose more resistance if they face stronger cost incentives. Hence, as A3 states, under any ownership structure whatsoever, as incentive intensity goes up, the insistence on compensation by the supply side will impose higher adaptation costs.

Bajari and Tadelis (2001) do not include governance in their analysis of contracts, so we need to consider ways in which changes in governance will change the marginal effect of incentives on adaptation costs. To fix ideas, begin with a fixed price contract, z = 1, where adaptation costs are highest, and consider a reduction in incentives (a decrease in z) that will in turn reduce adaptation costs. From A2, in a hierarchy the adaptation costs are lower than in market governance. From A3, a decrease in supplier incentives will reduce adaptation costs. It is reasonable therefore to assume that A2 and A3 together are consistent with improved

marginal reductions in adaptation costs when hierarchy is chosen. Formally we assume:

A4: Reducing adaptation costs by weakening incentives is more effective under hierarchy: $\frac{\partial k(z,H)}{\partial z} > \frac{\partial k(z,M)}{\partial z} > 0$.

Recall that our simplifying asymmetric approach to adaptation assumes that the supplier never initiates a need for adaptation. To keep things simple, we assume that when adaptation happens then we are not affecting the utility of the supplier (he is exactly compensated for any extra costs that adaptation imposes on him). As such, adaptation causes extra costs and inefficiencies that are imposed on the buyer or the interface-coordinator, but no extra rents are gained or lost by the supplier. This is not an important assumption but it conveniently simplifies the analysis.²⁷

4.4 Selective Intervention

It is important to make explicit what has been so far implicit. The choice of governance structure affects both production costs, c(e, G), and adaptation costs, k(z, G). Assumptions A1 through A4 imply that supplier autonomy defined by market governance reduces production costs but increases adaptation costs, while hierarchy does the opposite. There is, perhaps, an obvious solution: let the supplier retain control of production decisions related to the original design, and let the interface-coordinator (or possibly the buyer) retain control over adaptation de-

²⁷In Bajari and Tadelis (2001), when adaptation occurs then there is some expected rents to the supplier due to lock-in, and they are competed away *ex ante*. This could be included here with a more sophisticated focus on *ex post* adaptation with asymmetric information, but it would not shed light on the forces we wish to illuminate.

cisions. This, however, is ruled out by the assumption that *selective intervention* is severely limited:

A5: Administrative control is allocated over both production and adaptation, and the two processes cannot be separated to allow for selective intervention.

This is in line with the discussion in section 3.4 where it was argued that it is impossible to combine replication and selective intervention (Williamson, 1985, chap.6). The model takes an extreme position on the inability to have selective intervention, but clearly what matters is that at the margin some selective intervention is impossible. Realistically, a contract may specify some domains of control that each party may have. As long as contracts are incomplete, however, the impossibility of perfect selective intervention will persist.

4.5 Markets versus Hierarchies: Governance and Incentives

We now proceed with the analysis of endogenously choosing governance and incentives. First consider the objective of the supplier,

$$\max_{e \ge 0} u_S(e; z, G) = F - zc(e, G) - y(e)$$

The assumptions on c(e, G) and on y(e) imply that the the supplier's optimal choice $e^*(z, G)$ is increasing in z (stronger incentives increase effort), that $e^*(0, G) = 0$ (without incentives there is no effort) and for any $z \in (0, 1]$ A1 implies that $e^*(z, M) > e^*(z, H) > 0$ (effort is higher when the supplier controls production through market governance.)

Given the optimal response of the supplier, surplus maximization is given by the following program,²⁸

$$\max_{\substack{G \in \{H,M\}\\z \in [0,1]}} U(z,G;\tau,\sigma) = v \underbrace{-c(e^*(z,G),G) - y(e^*(z,G))}_{\text{Production Costs and Compensation}} \underbrace{-\sigma(1-\tau)k(z,G)}_{\text{Expected Adaptation Costs}}$$

In the appendix we show that this objective function exhibits increasing differences in all of its arguments, resulting in monotone comparative statics (see, e.g., Milgrom and Shannon, 1994) that in turn implies:

Proposition: When Asset specificity increases (higher σ), or when contracts are more incomplete (transactions are more complex - lower τ), the relative benefits of hierarchy over markets increase. Furthermore, optimal incentives become weaker.

The intuition for this result is quite straightforward. Strong cost-reducing incentives reduce production costs at the expense of raising adaptation costs (assumption A3). Market governance makes cost-reducing incentives more effective (assumption A1) at the expense of raising adaptation costs (assumption A2). Hence, more asset specificity and more contractual incompleteness, which increase the expected costs of adaptations, favor the use of hierarchies with weak cost-reducing incentives. Because at the margin, controlling adaptation costs through incentives is more effective under hierarchy (assumption A4), incentives and governance choices are *complements* in the organizational design. This result is not

²⁸We substitute $e^*(s,G)$ in place of e into the objective function to take account of the supplier's incentive compatibility constraint. We assume that the participation constraint does not impose a problem, which in itself is innocuous. The buyer's expected utility is $u_B(z,G;\tau) = v - F - (1-z)c(e^*(z,G),G) - (1-\tau)k(z,G)$, and total surplus is just $U(\cdot) = u_B(\cdot) + u_S(\cdot)$.

transparent in the less formal arguments of TCE. Finally, all this rests on the impossibility of selective intervention (assumption A5). Without it, one can assign control of different processes selectively to the agent who is most impacted, reducing adaptation costs dramatically without the sacrifice of lower productivity. Illuminating the complementarity between governance and incentives is a useful contribution of the formal model with respect to the semi-formal arguments in section 3.²⁹

It is useful to illustrate the result through a simple numerical example. Let $c(e,G) = \overline{c} - Ge$, k(z,G) = G + 4z and $y(e) = \frac{e^2}{2}$ where G = 1 represents hierarchy, and G = 2 represents market. Given an incentive scheme F + (1-z)c, the supplier chooses $e^*(z,G)$ to maximize $F - z(2-Ge) - \frac{e^2}{2}$ resulting in the optimal choice $e^*(z,G) = zG$. Substituting this into the objective function of surplus maximization yields,

$$\max_{\substack{z \in [0,1] \\ G \in \{1,2\}}} U = v - (\overline{c} - G^2 z) - \sigma (1 - \tau)(G + 4z) - \frac{z^2 G^2}{2}$$

Since $G \in \{1, 2\}$ corresponds to $\{H, M\}$, we can separately solve for $z_H^*(\sigma, \tau) = \max\{0, 1 - 4\sigma(1 - \tau)\}$ and $z_M^*(\sigma, \tau) = 1 - \sigma(1 - \tau)$. The resulting optimal surplus

²⁹The appropriate legal rules would strengthen our result. Namely, if the contract stipulates that the supplier retains control of supply processes, then the courts will enforce the letter of the contract that defines performance (design) and payments. This gives the supplier more bargaining power when adaptation is requested – he can use the original contract as a threat. This imposes extra costs due to haggling during the renegotiation process. If, however, the contract stipulates that both stages implement adaptations at the direction of the interface coordinator, then the courts will not enforce the performance measures under the doctrine that if the interface-coordinator maintains control and can direct the parties as it wishes, then the supplier (or buyer) cannot be made accountable for detailed design based performance. This is the situation where courts will defer to the power of fiat, as is common in employment relationships. As such, adaptation is further facilitated when hierarchy is selected.

for market is $U_M^*(\sigma,\tau) = v - \overline{c} - 6\sigma(1-\tau) + 2(\sigma(1-\tau))^2 + 2$ and for hierarchy is

$$U_H^*(\sigma,\tau) = \begin{cases} v - \overline{c} - 5\sigma(1-\tau) + 8(\sigma(1-\tau)^2 + \frac{1}{2} & \text{if } 0 \le \sigma(1-\tau) \le \frac{1}{4} \\ v - \overline{c} - \sigma(1-\tau) & \text{if } \frac{1}{4} \le \sigma(1-\tau) \le 1 \end{cases}$$

Comparing the expressions for total surplus we obtain that $U_H^*(\sigma,\tau) > U_M^*(\sigma,\tau)$ if and only if $\sigma(1-\tau) > 0.5$. For example, if the contract is not too incomplete $(\tau > 0.5)$ or if the transaction is not too specific $(\sigma < 0.5)$ then markets are always optimal. If, however, the contract is totally incomplete $(\tau = 0)$ then for $\sigma > 0.5$ hierarchy will dominate market because it has lower total transaction costs. This corresponds exactly to figure 2 with $\sigma^* = 0.5$.

In reference to Figure 2, as contractual incompleteness increases, total transaction costs increase for both markets and hierarchies, causing both functions to rise. However, market transactions costs increase faster, making hierarchies better for a larger range of specificity σ . This is a point that the model illuminates. Since σ and $(1-\tau)$ are multiplied, contractual incompleteness and asset specificity are complementary. This in turn implies that the negative effects that incompleteness has on adaptation costs are exacerbated when asset specificity is higher. One can consider other specifications where this would not be the case, yet interestingly, this specification is consistent with the empirical results of Masten (1984).

5 Extensions and Applications

5.1 Hybrids

The advantages of the two polar modes, simple market exchange and hierarchy, are pronounced for transactions that deviate significantly from the crossover value σ^* (as shown in Figure 2). "Hybrid" contracting arrangements, however, that are located between hierarchy and simple spot market transactions, appear to be

common and thus warrant brief consideration. Taking adaptation as the central problem of economic organization, the mechanisms in hybrid arrangements that operate in the support of semi-cooperative adaptations need to be identified. We argue that when transaction are in the intermediate range of hazards (close to the value σ^*), the main *mechanism* that supports coordinated adaptation without imposing the burdens of hierarchy is the use of *credible commitment*.

In some instances credible commitments are contractually agreed upon by the parties to help provide additional assurances, such as the thirty-two year coal supply agreement between the Nevada Power Company and the Northwest Trading Company is illustrative (Williamson, 1991, pp. 272-273). These may include additional supports such as formal auditing and information disclosure procedures, as well as reliance on private ordering dispute resolution mechanisms that are outside the formal courts (Llewellyn, 1931; Macaulay, 1963; Summers, 1969; Macneil, 1974, Galanter, 1981). Another example is the use of "hostages to support exchange" (Williamson, 1983) where defection from cooperation is deterred by reciprocal exposure of specific assets that experience positively correlated disturbances.

5.2 Reality Testing

TCE addresses itself to issues of scaling up, remediableness, and bureaucracy and recommends that other theories of firm and market organization consider doing likewise.

Scaling up. The object of a simple model is to capture the essence, thereby to explain puzzling practices and make predictions that are subjected to empirical testing. But that is not the only relevant test. Simple models can also be "tested" with respect to scaling up. Does repeated application of the basic mechanism out

of which the simple model works yield a result that recognizably describes the phenomenon in question?

The test of scaling up is often ignored, possibly out of awareness that scaling up cannot be done. Sometimes it is scanted, possibly in the belief that scaling up can be accomplished easily. We advise that claims of real world relevance, including public policy relevance, of any candidate theory of the firm that cannot be shown to scale up from toy model status to approximate the phenomenon of interest (usually, the modern corporation) should be regarded with caution.

With reference to the theory of the firm as governance structure the question is this: Does successive application of the make-or-buy decision, as it is applied to individual transactions in the TCE setup, scale up to describe something that approximates a multi-stage firm? Note that, as described previously, TCE assumes that the transactions of principal interest are those that take place at the interface between (rather than within) technologically separable stages, which is the "boundary of the firm" issue as described elsewhere (Williamson, 1985, p. 96-98). Upon taking the technological "core" as given, attention is focused as a series of separable make-or-buy decisions - backward, forward, and lateral - to ascertain which should be outsourced and which should be incorporated within the ownership boundary of the firm. So described, the firm is the inclusive set of transactions for which the decision is to make rather than buy - which does implement scaling up, or at least is an approximation thereto (Williamson, 1985, pp. 96-98).

Remediableness. Public policy analysts have often assumed that transaction

³⁰In the context of a multi-divisional firm (Chandler, 1962) the scaling up to which we refer corresponds more nearly to that of an operating division (which could be a freestanding firm that has been acquired by a diversified enterprise) than the entire multidivisional enterprise – where the "general office" provides an additional level of strategic participation.

costs in the public sector are zero. That is unrealistic, yet standard public policy proceeded in an asymmetric way: private sector contracting experienced market failures, by reason of positive transaction costs, but there was no corresponding concept for public sector failures.³¹ Little surprise, then, that convoluted public policy prescriptions were often (unwittingly) anchored in asymmetric application of zero transaction cost reasoning, of which regulation is an example (Coase, 1964).

The remediableness criterion is an effort to deal symmetrically with real world institutions, both public and private. The criterion is this: an existing mode of organization for which no superior feasible form of organization can be described and implemented with expected net gains is presumed to be efficient (Williamson, 1996, Chap. 8). In other words, a "revealed preference" approach is applied to the choice of governance.

Because all feasible modes of organization are flawed, the feasibility stipulation precludes all appeals to the fiction of zero transaction costs (in any sector whatsoever – public, private, nonprofit, etc.) from the very outset. The implementation stipulation requires that the costs of implementing a proposed feasible alternative (one that is judged to be superior to an extant mode in a de novo side-by-side comparison) be included in the net gain calculus.³²

The upshot is that the remediableness criterion is an effort to disallow asymmetric efficiency reasoning of a zero transaction cost kind and force the relevant

³¹Albeit a caricature, "normative public policy analysis began by supposing that ... policy was made by an omnipotent, omniscient, and benevolent dictator" (Dixit, 1996, p. 8) – which, in transaction cost terms, assumes the absence of implementation obstacles, bounds on rationality, and opportunism, respectively.

³²The presumption that a chosen mode is efficient if the expected net gain is negative is nevertheless rebuttable: it can be challenged by showing that the obstacles to implementing an otherwise superior feasible alternative are "unfair." See Williamson (1996) for a discussion.

efficiency issues for the making of public policy – namely, feasibility, implementation, and rebuttal – to the surface.

Burdens of Bureaucracy. The impossibility of implementing either replication or selective intervention is partly attributable to the incentive distortions that arise when successive stages of production that had previously been independent are placed under unified ownership (to include the control over accounting practices that accrues thereto). Upon recognizing that the hypothetical gains of combining replication with selective intervention cannot be realized in practice, such efforts will not be undertaken. Unavoidable trade-offs are posed wherein incentive intensity is reduced in the service of coordinated adaptation when transactions are taken out of the market and organized internally – as reflected in the discrete structural differences between markets and hierarchies (as displayed in Table 1 and derived in Section 4).

But there is more to it than that. Reaching beyond the scope of the analysis in sections 3 and 4, some of the more insidious burdens of bureaucracy arise because "integration affects the internal politics of the corporation with systematic performance consequences" (Williamson, 1985, p. 145). These include tilting the managerial promotion game away from merit in favor of politics (sub-goal pursuit), the increased propensity to intervene (over-manage and over-control), limits upon internal incentives due to internal equity considerations, and operating and investment decisions are distorted (Williamson, 1985, pp. 147-152). The appearance of what Paul Milgrom and John Roberts refer to as "influence costs" (1988) are also in the bureaucratic burden spirit.

Our understanding of the strengths and weaknesses of bureaucracy is unfortunately very underdeveloped in comparison with our understanding of the strengths and weaknesses of markets – mainly because bureaucracy is both a comparatively neglected and a formidably difficult subject. A more informed assessment of markets and hierarchies requires that this condition be corrected by inviting future research.

5.3 Variations on a Theme

Many of the regularities that are associated with the intermediate product market transaction recur, in variable degree, as variations on a theme. Transaction cost economics not only has many applications within the field of industrial organization but within most applied fields of economics as well – to include labor, public finance, comparative economic systems, and economic development and reform. Applications to business – to the fields of strategy, organizational behavior, marketing, finance, operations management, and accounting – are likewise numerous. Numerous applications to the contiguous social sciences (especially sociology, political science, social psychology, and aspects of the law) have also been made. More generally, transaction cost economics has broad reach because any problem that arises as or can be reformulated as a contracting problem can be examined to advantage in transaction cost economics terms. [****cites will be added – in text or footnote****]

5.4 Empirical Evidence

Shortly after the main ideas of TCE were laid out, several seminal empirical papers confirmed the main prediction of the theory, including Monteverde and Teece (1982), Masten (1984) and Joskow (1985). Jeffrey Macher and Barak Richman (2008) report that there were over 900 published empirical papers on TCE over the

period 1980 to 2004, with more in progress. Indeed, "despite what almost 30 years ago may have appeared to be insurmountable obstacles to acquiring the relevant data [which are often primary data of a microanalytic kind], today transaction cost economics stands on a remarkably broad empirical foundation" (Geyskens, Steenkamp and Kumar 2006). As Michael Whinston puts it: TCE has been "one of the great [empirical] success stories in industrial organization over the past [30] years" (Whinston, 2001, p. 185). There is no gainsaying that TCE has been much more influential because of its broad and varied applications and the empirical work that it has engendered. Most recently, new empirical studies have focused attention on the importance of adaptation to empirically validate the central themes of TCE (e.g., Silke Forbes and Mara Lederman, 2010, and Arnaud Costinot, Lindsay Oldenski and James Rauch, 2010).

5.5 Antitrust and Regulation

TCE was stimulated in part by the disarray in antitrust enforcement and regulation during the 1960s. As Coase observed, "If an economist finds something – a business practice of one sort or another that he does not understand, he looks for a monopoly explanation. And as in this field we are very ignorant, the number of ununderstandable practices tends to be very large, and the reliance on a monopoly explanation, frequent" (1972, p. 67). The possibility that complex contract and organization had beneficial governance purposes was ignored. The compound implausibility of such monopoly reasoning coupled with a small but growing interest in transaction cost economizing would eventually help to bring relief from such excesses.³³

³³Interestingly, Timothy Muris, during his term as chair of the Federal Trade Commission, held that much of the New Institutional Economics "literature has significant potential to improve

As applied to regulation and deregulation, more attention to the details and to strategic behavior on the part of participants in regulated/deregulated sectors needs to be factored in.³⁴ As Paul Joskow observes of efforts to deregulate electricity in California, too much deference was given to the (assumed) efficacy of smoothly functioning markets and insufficient attention to potential investment and contractual hazards and appropriate governance responses thereto: "Many policy makers and fellow travelers have been surprised by how difficult it has been to create wholesale electricity markets... Had policy makers viewed the restructuring challenge using a TCE framework, these potential problems are more likely to have been identified and mechanisms adopted ex ante to fix them" (2000, p. 51).

6 Some Challenges

Although efforts have been made to develop fully formal models of transaction cost economics, such efforts remain a work-in-progress. As Williamson (2010) explains, transaction cost economics, like many other theories, has undergone a natural antitrust analysis and policy. In particular, ... [the transaction cost branch has] focused on demystifying the 'black box' firm and on clarifying important determinants of vertical relationships" (2003, p. 15). Opening the black box and acquiring an understanding of the mechanisms inside has had an impact, moreover, on practice (Muris, 2003, p. 11; emphasis in original): "The most impressive recent competition policy work I have seen reflects the NIE's teachings about the appropriate approach to antitrust analysis. Much of the FTC's best work follows the tenets of the NIE and reflects careful, fact-based analyses that properly account for institutions and all relevant theories, not just market structures and [monopoly] power theories."

³⁴Interestingly, the first "empirical" application of TCE was to regulation. The issue was the purported efficacy of franchise bidding for natural monopoly, as advanced by Demsetz (1968) and applied to the cable television industry by Richard Posner (1972). Missing from both of these sanguine assessments of the efficacy of franchise bidding were (1) an examination of the details by which franchise bidding would be implemented and (2) an awareness that the nonredeployability of asset specific investments would pose serious (but unexamined) problems (Williamson, 1976).

progression from informal (1930-1970), to pre-formal (1970s), to semi-formal (1980s and beyond), of which full formalization is the last step.

The path breaking paper by Sanford Grossman and Oliver Hart (1986) and the follow-on paper by Hart and John Moore (1990), which founded the "property-rights theory" (PRT) literature, have been very influential. They have established the modeling apparatus used in section 4 above in which governance choices affect the incentives of parties to a transaction. That said, there are three fundamental differences between the PRT and TCE approaches.

First, while transaction cost economics emphasizes ex post adaptation as the main problem of organization, the property rights approach has focused more on the problem of ex ante alignment of incentives to invest in specific assets, mainly of the intangible (human asset) kind. This same focus on ex ante incentives is also at the core of most agency-based theories.

Second, PRT and agency based theories are silent about how transactions differ with respect to complexity, uncertainty and contractual incompleteness. Instead, agency theory and property rights theory explore the consequences of incomplete information in that some aspects of the transaction have zero costs of contracting while others have infinite costs. Namely, in standard agency models, output measures of the transaction (say the component's functionality, quality, or a signal of these) can be specified at no cost, while the effort and actions of people who work towards achieving the desired goals cannot be specified at any cost. The PRT paradigm assumes that contingent decisions related to how to do the job or how to respond to changes in the environment cannot be specified at any cost, while it is costless to decide who has the rights to make decisions and execute them, and it is costless to enforce these rights. TCE, by contrast, focuses on measur-

able dimensions over which transactions differ, with emphasis on identifying how different kinds of transactions are discriminately allocated to different governance structures. This accounts for much of its predictive content and empirical success described in section 5.4.

Third, PRT identifies ownership of assets as the critical variable, and takes the view that ownership determines who has the residual right to decide how to use the productive assets. As such, changes in ownership of assets from one party to another will change the incentives of the parties to invest in the specific relationship, implying a unified framework for the costs and benefits of integration. The implications are that separable, well defined assets ought to be owned by the individuals whose incentives are most affected by ownership. But there is more: directional ownership is predicted. PRT asserts that either buyer or seller should own the assets, and it matters who the owner is. In fact, however, bundles of assets are owned by firms, not by individuals. Holmstrom (1999), who raises this important critique, argues that owning bundles of assets allows the firm to internalize many of the externalities that are associated with incentive designs in a world characterized by informational imperfections, which are responsible for contractual incompleteness. By associating the decision maker with more assets under his control, his administrative control is expanded, and he possesses more levers through which to implement coordinated adaptation. This observation is very much in line with our notion of the interface coordinator being the executive who has the authority to coordinate adaptation as needed, even though he as an individual need not own the assets.

Still, the Grossman-Hart methodology is an instrumental contribution. Building on their methods, and the insights in Bajari and Tadelis (2001), Tadelis (2002)

makes a step towards formalizing some of the central ideas of TCE. The model we offer in section 4 is an attempt to push the full formalization of TCE yet another modest step forward. More work, however, still needs to be done. A full incorporation of the role of law in formal models is very challenging.

7 Conclusions

• • •

References

- Alchian, Armen, and H. Demsetz. 1972. "Production, Information Costs, and Economic Organization," American Economic Review, 62 (December): 777-795.
- Ashenfelter, Orley and David Bloom (1993) "Lawyers as Agents of the Devil in a Prisoner's Dilemma Game," NBER working paper no. 4447
- Bajari, Patrick, and Steven Tadelis. 2001. "Incentives Versus Transaction Costs: A Theory of Procurement Contracts" Rand Journal of Economics 32(3):387-407
- Baron, James N., and David M. Kreps. 1999. Strategic Human Resources: Frameworks for General Managers. New York: John Wiley.
- Barnard, Chester. 1938. The Functions of the Executive. Cambridge: Harvard University Press (fifteenth printing, 1962).
- Boerner, Christopher S., and Jeffrey T. Macher. 2000. "Transaction Cost Economics: A Review and Assessment of the Empirical Literature,"
- Chandler, Alfred D. 1966. Strategy and Structure. New York: Doubleday & Co.
- Coase, Ronald H. 1937. "The Nature of the Firm," Economica N.S., 4: 386-405. Reprinted in Oliver E. Williamson and Sidney Winter, eds., 1991. The Nature of the Firm: Origins, Evolution, Development. New York: Oxford University Press, pp. 18-33.
- Coase, Ronald H. 1972. "Industrial Organization: A Proposal for Research," in V. R. Fuchs, ed., Policy Issues and Research Opportunities in Industrial Organization. New York: National Bureau of Economic Research, pp. 59-73.

- Commons, John R. 1932. "The Problem of Correlating Law, Economics, and Ethics," Wisconsin Law Review, 8: 3-26.
- Crocker, Keith, and Scott Masten. 1996. "Regulation and Administered Contracts Revisited: Lessons from Transaction-Cost Economics for Public Utility Regulation," Journal of Regulatory Economics, 8: 5-39.
- Cyert, Richard M., and James G. March. 1963. A Behavioral Theory of the Firm. Englewood Cliffs, NJ: Prentice-Hall.
- Demsetz, Harold. 1983. "The Structure of Ownership and the Theory of the Firm/Comment," Journal of Law & Economics, 26 (June): 375-393.
- Dixit, Avinash. 1996. The Making of Economic Policy: A Transaction Cost Politics Perspective. Cambridge, MA:MIT Press.
- Fama, Eugene F., and Michael C. Jensen. 1983. "Separation of Ownership and Control," Journal of Law and Economics, 26 (June): 301-326.
- Gibbons, Robert. 2000. "Why Organizations are Such a Mess (and What an Economist Might Do About It),

Gibbons JEBO 2005

Grossman and Hart (1986)

Hart (1995)

Hart and Holmstrom (2004)

Hart and Moore (1990)

Holmstrom (1999)

Holmstrom and Milgrom (1991)

- Hayek, Friedrich. 1945. "The Use of Knowledge in Society," American Economic Review, 35 (September): 519-530.
- Joskow, Paul. 1991. "The Role of Transaction Cost Economics in Antitrust and Public Utility Regulatory Policies," Journal of Law, Economics, and Organization, 7 (Special Issue): 53-83.
- Joskow, Paul L. 2000. "Transaction Cost Economics and Competition Policy," unpublished manuscript.
- Klein, Benjamin. 1980. "Transaction Cost Determinants of 'Unfair Contractual Arrangements," American Economic Review, 70 (May): 356-362.

- Kreps, David M. 1990. "Corporate Culture and Economic Theory," in James Alt and Kenneth Shepsle, eds., Perspectives on Positive Political Economy. New York: Cambridge University Press, pp. 90-143.
- Kreps, David M. 1999. "Markets and Hierarchies and (Mathematical) Economic Theory," in Glenn Carroll and David Teece, eds., Firms, Markets, and Hierarchies. New York: Oxford University Press, 121-155.
- Levin Jonathan and Steven Tadelis (2010) "Contracting for Government Services: Theory and Evidence from U.S. Cities," Journal of Industrial Economics 58(3):507-541
- Levy, Brian, and Pablo Spiller. 1994. "The Institutional Foundations of Regulatory Commitment: A Comparative Analysis of Telecommunications Regulation," Journal of Law, Economics and Organization, 10 (October): 201-246.
- Levy, Brian, and Pablo Spiller. 1996. Regulations, Institutions, and Commitment: Comparative Studies of Telecommunications. Cambridge University Press.
- Lyons, Bruce R. 1996. "Empirical Relevance of Efficient Contract Theory: Inter-Firm Contracts," Oxford Review of Economic Policy, 12 (No. 4): 27-52.
- Macneil, Ian R. 1974. "The Many Futures of Contracts," Southern California Law Review, 47 (May): 691-816.
- March, James G., and Herbert A. Simon. 1958. Organizations. New York: John Wiley & Sons.
- Masten, Scott, and Stephane Saussier. 2000. "Econometrics of Contracts: An Assessment of Developments in the Empirical Literature on Contracting," Revue D'Economie Industrielle, 92: 215-236.

Milgrom and Roberts (1990)

Milgrom and Roberts (1992)

- North, Douglass. 1990. "A Transaction Cost Theory of Politics," Journal of Theoretical Politics, 2 (No. 4): 355-367.
- Posner, Richard A. 1972. "The Appropriate Scope of Regulation in the Cable Television Industry," The Bell Journal of Economics and Management Science, 3, No. 1 (Spring): 98-129.
- Posner, Richard A. 1993. "The New Institutional Economics Meets Law and Economics," Journal of Institutional and Theoretical Economics, 149 (March): 73-87.

- Riordan, Michael, and Oliver Williamson. 1985. "Asset Specificity and Economic Organization," International Journal of Industrial Organization, 3: 365-378.
- Selznick, Philip. 1950. "The Iron Law of Bureaucracy," Modern Review, 3: 157-165
- Shelanski, Howard, and Peter Klein. 1995. "Empirical Research in Transaction Cost Economics: A Review and Assessment," Journal of Law, Economics, and Organization, 11 (October): 335-361.
- Simon, Herbert. 1951.
- Simon, Herbert. 1957b. Models of Man. New York: John Wiley & Sons.
- Skaperdas, Stergios (1992) "Cooperation, Conflict, and Power in the Absence of Property Rights," American Economic Review, 82(4):720-39
- Tadelis, Steven (2002) "Complexity, Flexibility and the Make-or-Buy Decision," American Economic Review Papers and Proceedings 92(2):433-437
- Whinston, Michael (2000) "On the Transaction Cost Determinants of Vertical Integration," The Journal of Law, Economics, and Organization, Vol. 19, pp. 1-23, 2003
- Williamson, Oliver E. 1975. Markets and Hierarchies: Analysis and Antitrust Implications. New York: Free Press.
- Williamson, Oliver E. 1983. "Credible Commitments: Using Hostages To Support Exchange," American Economic Review, 73 (September): 519-540.
- Williamson, Oliver E. 1985. The Economic Institutions of Capitalism. New York: Free Press.
- Williamson, Oliver E. 1991. "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives," Administrative Science Quarterly, 36 (June): 269-296.
- Williamson, Oliver E. 1993. "Transaction Cost Economics and Organization Theory," Institutional and Corporate Change, 2 (2): 107-156.
- Williamson, Oliver E. 1996. The Mechanisms of Governance. New York: Oxford University Press.
- Williamson, Oliver E. 1999. "Public and Private Bureaucracies," Journal of Law, Economics, and Organization, 15 (April): 306-342.
- Williamson, Oliver E. 2000.

Williamson, Oliver E., Michael L. Wachter, and Jeffrey E. Harris.1975. "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange," Bell Journal of Economics, 6 (Spring): 25