The research program on which I and others have been working has been variously described as the “economics of governance,” the “economics of organization,” and “transaction cost economics.” As discussed in Section I, governance is the overarching concept and transaction cost economics is the means by which to breathe operational content into governance and organization. The specific issue that drew me into this research project was the puzzle posed by Ronald Coase in 1937: What efficiency factors determine when a firm produces a good or service to its own needs rather than outsource? As described in Section II, my 1971 paper on “The Vertical Integration of Production” made headway with this issue and invited follow-on research that would eventually come to be referred to as transaction cost economics. The rudiments of transaction cost economics are set out in Section III. Puzzles and challenges that arose and would require “pushing the logic of efficient governance to completion” are examined briefly in Section IV. Concluding remarks follow.

I. An Overview

For economists, if not more generally, governance and organization are important if and as these are made susceptible to analysis. As described here, breathing operational content into the concept of governance would entail examining economic organization through the lens of contract (rather than the neoclassical lens of choice), recognizing that this was an interdisciplinary project where economics and organization theory (and, later, aspects of the law) were joined, and introducing hitherto neglected transaction costs into the analysis. A predictive theory of economic organization was the object. The puzzle of vertical integration was an obvious place to start.

A. Governance

Whereas textbook microeconomic theory was silent on the concept of good governance, John R. Commons, who was a leading institutional economist during the first half of the twentieth century, formulated the problem of economic organization as follows: “The ultimate unit of activity … must contain in itself the three principles of conflict, mutuality, and order. This unit is a transaction” (Commons 1932, 4). Commons thereafter recommended that “theories of economics center on transactions and working rules, on problems of organization, and on the … [ways] the organization of activity is … stabilized” (Commons 1950, 21).
This conception of economics is to be contrasted with the neoclassical resource allocation paradigm in two important respects: first, whereas Commons viewed organization and the continuity of contractual relations as important, the resource allocation paradigm made negligible provision for either but focused instead on prices and output, supply and demand; second, whereas the price theoretic approach to economics would become the “dominant paradigm” during the twentieth century (Melvin W. Reder 1999, 43), institutional economics 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 (Stigler as reported in Edmund W. Kitch 1983, 170). Stalwarts notwithstanding, institutional economics “ran itself into the sand.”

This does not imply, however, that institutional economics was lacking for good ideas. Indeed, the Commons Triple of conflict, mutuality, and order prefigures the concept of governance as herein employed—in that governance is the means by which to infuse order, thereby to mitigate conflict and realize mutual gain. Furthermore, the transaction is made the basic unit of analysis.

James M. Buchanan (1975, 225) subsequently 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. If “mutuality of advantage from voluntary exchange is … the most fundamental of all understandings in economics” (Buchanan 2001, 29), then the lens of contract approach is an underused perspective.

The past 35 years have witnessed growing interest in the use of the lens of contract, to include both theories that emphasize ex ante incentive alignment (agency theory/mechanism-design, team theory, property rights theory) and those for which the ex post governance of contractual relations is where the main analytical action resides. Transaction cost economics is an ex post governance construction, with emphasis on those transactions to which Commons called attention—namely those for which continuity (or breakdown) of the exchange relation is of special importance. How did the attributes of such transactions differ from the ideal transaction, in both law and economics, of simple market exchange (where no such continuity relation was implied)? What were the governance ramifications?

Answers to these queries would entail reformulating the problem of economic organization in comparative contractual terms by (i) naming the key attributes with respect to which transactions differ, (ii) describing the clusters of attributes that define alternative modes of governance (of which markets and hierarchies are two), (iii) joining these parts by appealing to the efficient alignment hypothesis, wherein (iv) predictions would be derived to which empirical tests would be applied, and (v) public policy ramifications would be worked up. Antecedent to the foregoing, the contract relevant attributes of human actors would be named and explicated.

B. Organization

 Whereas the neoclassical theory of the firm treated the firm as a black box for transforming inputs into outputs according to the laws of technology, this was not, as Harold Demsetz (1983, 377) observed, an all-purpose construction. It is a “mistake to confuse the firm of [neoclassical] economic theory with its real-world namesake. The chief mission of neoclassical economics is to understand how the price system coordinates the use of resources, not the inner workings of real firms.”

Although Demsetz did not make the case that economics and organization theory should be joined in a combined effort to understand firm and market organization of a real world kind, that is nevertheless the research need and opportunity as I perceived it—in no small measure because of my training (1960–1963) in the PhD program at the Graduate School of Industrial Administration, Carnegie Mellon University. This remarkable program in interdisciplinary
social science made the case that organization theory should both inform and be informed by economics. Herbert Simon, James March, and Richard Cyert played especially important roles in putting this across. Considerations of bounded rationality, the specification of goals, intertemporal regularities (wherein organization takes on “a life of its own”), the critical importance of adaptation, the reliance within the operating parts on routines, and, more generally, the “architecture of complexity” were all basic concepts that would prove to be pertinent to an understanding of incomplete contracting and complex organization. Had the governance of contractual relations come under study at Carnegie Mellon, there is no question that this would have been examined in an interdisciplinary way.

C. Transaction Costs

Ronald Coase, in his classic 1937 paper on “The Nature of the Firm,” was the first to bring the concept of transaction costs to bear on the study of firm and market organization. The youthful Coase (then 27 years old) uncovered 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, 388), Coase observed that the decision to use one mode rather than the other should not be taken as given (as was the prevailing practice) but should be derived. Accordingly, Coase (1937, 389) 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 that allocation is dependent on the entrepreneur-coordinator. We have to explain the basis on which, in practice, this choice between alternatives is effected.

The missing concept was “transaction cost.”

The lapse to which Coase referred had little immediate effect (Coase 1988, 23) and failed to take hold over the next 20 years, during which period the implicit assumption of zero transaction costs 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 implications of this standard assumption were displayed for all to see.

The first demonstration was Coase’s 1960 article on “The Problem of Social Cost.” Upon reformulating the externality problem in contractual terms and pushing the logic of zero transaction cost reasoning to completion, he realized 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, 717). That is because if transaction costs are

1 Jacques Dreze speaks for me, and, I am sure, for many others in his statement that “Never since [my visit to Carnegie Mellon] have I experienced such intellectual excitement” (1995, 123). Nobel Laureates in economics from the small group of faculty and students at CMU include Herbert Simon, Franco Modigliani, Merton Miller, Robert Lucas, Edward Prescott, and Finn Kydland.

2 Classic books by Carnegie Mellon faculty that feature economics and organization theory include Models of Man (Simon 1957b), Organizations (March and Simon 1958), and the Behavioral Theory of the Firm (Cyert and March 1963).

3 One way to introduce organizational considerations is to change the objective function of the firm by supplanting the neoclassical assumption of profit maximization with various forms of “managerial discretion”—such as sales maximization (Baumol 1959), growth maximization (Marris 1964), or expense preference (Williamson 1964). These efforts to introduce “realism in motivation” yielded few predictions and resulted in little empirical testing.

4 Even the Chicago School, which had grave reservations with overreaching uses of externality arguments, was resistant to Coase’s claims that externalities vanished under zero transaction cost conditions. For a discussion of Coase versus Chicago, see Edmund Kitch (1983, 220–221).
zero then the parties to tort transactions will costlessly bargain to an efficient result whichever way property rights are assigned at the outset. In that event, the emperor really did have no clothes: externalities and frictions of other kinds would vanish. That being preposterous, the real message was this: “study the world of positive transaction costs” (Coase 1992, 717). Kenneth J. 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” (Arrow 1969, 48).

But whereas pushing the logic of zero transaction costs to completion would reveal the need to make provision for positive transaction costs, there were 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. That would prove to be a curse, in that some form of transaction cost could be invoked to explain any condition whatsoever after the fact, as a result of which appeal to transaction costs acquired a “well deserved bad name” (Stanley Fischer 1977, 322). Second, it does not suffice to show that some types of transaction costs are demonstrably great. Unless these costs differ among modes (say, as between markets and hierarchies), such a demonstration lacks comparative contractual significance. 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.

II. The Vertical Integration of Production

What I have referred to as the “Carnegie Triple” (Williamson 1996, 25) is this: be disciplined; be interdisciplinary; have an active mind. Being disciplined means to take your core discipline seriously and work at it on its own terms. Being interdisciplinary means to appeal to the contiguous social sciences—if and as the phenomena under study cross disciplinary lines. Having an active mind entails asking the question, “What is going on here?” rather than pronouncing “This is the law here.” The Carnegie Triple would serve me well when I named industrial organization as my field, even though I had never taken an industrial organization course at Carnegie Mellon (or elsewhere), and I went on the job market.

Coase (1972, 62) described the leading industrial organization textbooks in the 1960s as “applied price theory”—with which I agree, but with a caveat: the structure-conduct-performance paradigm also played an important role in the “Harvard School” approach to the field. The organization of markets (especially with respect to the number and size distribution of firms and the condition of entry) thus figured prominently, but the organization of firms was ignored. Because firms were production functions for transforming inputs into outputs according to the laws of technology, the IO public policy lesson was this: except as contracting practices and organizational structures had a physical or technical basis, nonstandard and unfamiliar

---

5 Not everyone agrees. Some economists take the “Coase Theorem” (the first 15 pages of Coase 1960) to imply that costless renegotiation accurately describes contracting in practice. However, the following 29 pages in Coase (1960) reveal that the zero transaction cost assumption is not only wrong but undermines our understanding of complex economic phenomena. Express provision for positive transaction costs would thereafter need to be made if externalities and other real world contractual phenomena are to be accurately understood. Coase reaffirmed that this was his purpose in his Nobel Prize Lecture (Coase 1992, 712).

6 For a discussion of these distinctions, see Roy D’Andrade (1986).
forms of contract and organization were regarded as deeply problematic and presumptively anticompetitive.7

By contrast with this one-sided interpretation of deviations from the norm under the prevailing IO orientation, the Carnegie Mellon perspective on contractual and organizational variety was that such could also serve efficiency purposes. This difference in perspectives would forcefully register when I served in 1966–67 as the Special Economic Assistant to the Head of the Antitrust Division of the US Department of Justice, especially when I was asked to comment on an early draft of the Schwinn brief. The issue was one of vertical market restrictions, and the brief advanced the argument that the franchise restrictions imposed by the bicycle manufacturer Schwinn on its (nonexclusive) franchisees were anticompetitive. My view was more cautious. Not only was it unclear to me that the restrictions had anticompetitive effects, but a case could be made that the restrictions in question served the purpose of preserving the integrity of the franchise system—additionally or instead (Williamson 1985, 183–189). Alas, the principal architects of the Schwinn brief viewed the case as an opportunity to apply the “then prevailing thinking of the economics profession on restricted distribution” (Richard A. Posner 1977, 1). This anticompetitive interpretation succeeded in arguments before the US Supreme Court.8

By reason of what I perceived to be truncated and defective reasoning with Schwinn and other cases,9 I decided to revisit vertical integration and vertical market restrictions when I resumed teaching at the University of Pennsylvania. The graduate students and I worked our way through the literature and, some very good papers notwithstanding (William Fellner 1947; Lionel W. McKenzie 1951; George J. Stigler 1951), satisfied ourselves that organizational economies played no significant role. I therefore decided to revisit vertical integration from a combined economics and organization theory perspective.

My paper on “The Vertical Integration of Production: Market Failure Considerations” differed from orthodoxy in the following respects: (i) I examined economic organization through the lens of contract (rather than orthodox lens of choice), (ii) described cognition in terms of bounded rationality, on which account all complex contracts are incomplete, (iii) made provision for strategic behavior (defection from the spirit of cooperation) when an outsourced good or service experienced disturbances for which the stakes are great, (iv) treated adaptation as the main efficiency purpose of economic organization, and (v) distinguished between investments in generic assets and specific assets, where a bilateral dependency relation between supplier and buyer stages was ascribed to the latter. Taken together, the argument comes down to this: efficient intermediate product market exchange is usually well served by simple market contracting if the assets are generic; but the advantage shifts to hierarchy as bilateral dependency (and the resulting risk of costly maladaptations) builds up by reason of asset specificity and outlier disturbances.

Although I initially regarded this paper as a standalone effort to solve the puzzle of the boundaries of the firm and expand our understanding of vertical integration, it turned out that vertical integration would become a paradigm for the study of complex contract and economic organization. The combination of incomplete contracts, bilateral dependency (contingent on asset specificity), and defection from the norm of coordinated adaptation when a contract experiences

---

7 Coase (1972, 67) described the prevailing monopoly predilection as follows: One important result of this preoccupation with the monopoly problem is that if an economist finds something—a business practice of one sort or other—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 rather large, and the reliance on a monopoly explanation, frequent.

8 Interestingly, the Supreme Court effectively reversed Schwinn a decade later as the limits of “prevailing thinking” became increasingly clear.

9 I do not, however, mean to suggest that my disagreements were common. The leadership and staff in the Antitrust Division in the late 1960s were superlative.
significant disturbances (for which the stakes are great) had application to a wide range of phenomena that were interpreted as variations on a theme.

The initial trick was to think contractually, which for many phenomena was easy but for others required that the phenomenon be reformulated in contracting terms. This, however, was merely the first step. The key concepts needed to be operationalized; a predictive theory needed to be worked up; and, as gaps and omissions arose, the logic of positive transaction costs would need to be pushed to completion. The first two are addressed in Section III and the last in Section IV.

III. The Rudiments

Upon realizing that this approach to the study of economic organization had broad application, I needed to work up the basic mechanisms and the underlying logic more systematically. The rudiments are described in three clusters: key conceptual moves; key operational moves; and applications. Common to all three clusters is the need to examine economic organization at a more microanalytic level of analysis, which is consistent with Simon’s observation that:

In the physical sciences, when errors of measurement and other noise are found to be of the same order of magnitude as the phenomena under study, the response is not to try to squeeze more information out of the data by statistical means; it is instead to find techniques for observing the phenomena at a higher level of resolution. The corresponding strategy for economics is obvious: to secure new kinds of data at the micro level. (Simon, 1984, 40)

What follows is a very compact summary.

A. Conceptual Moves

The basic moves here are to elaborate upon (i) the attributes of human actors and (ii) adaptation and to introduce (iii) contract laws (plural).

Human Actors.—If “Nothing is more fundamental in setting our research agenda and informing our research methods than our view of the nature of the human beings whose behavior we are studying” (Simon 1985, 303), then social scientists are challenged to name the cognitive, self-interest, and other attributes of human actors on which their analyses rest.

The cognitive assumption that Simon (1991, 1957) has characterized as his “lodestar” is bounded rationality, which he describes as behavior that is “intendedly rational, but only limitedly so.” Human actors, so described, are neither hyperrational nor irrational but are attempting effectively to cope with complex contracts that are incomplete.

Incompleteness notwithstanding, transaction cost economics also makes provision for “feasible foresight,” as reflected in George P. Shultz’s remarks about how his “training in economics has had a major influence on the way I think about public policy tasks, even when they have no particular relation to economics. Our discipline makes one think ahead, ask about indirect consequences, take note of variables not directly under consideration” (Shultz 1995, 1). This is a recurrent theme in the discussion in Section IV of pushing the logic to completion. I merely observe here that many economists and others within the social science community (Michels 1962; March and Simon 1958) practice feasible foresight, although this is an underused perspective.

My interpretation of Simon’s description of self-interest seeking as “frailty of motive” (Simon 1985, 303) is that most people will do what they say (and some will do more) most of the time
without self-consciously asking whether the effort is justified by expected discounted net gains. If they slip, it is a normal friction and often a matter of bemusement. The proposition that routines describe the behavior of most individuals most of the time contemplates (nonstrategic) benign behavior.

But while accurate descriptions of what is going on “most of the time” are plainly essential, much of what is interesting about human behavior in general and organizations in particular has reference not to routines but to exceptions. Indeed, once good routines have been developed, the chief role of management is to deal with exceptions. In the context of outsourcing, such exceptions arise from contractual incompleteness in combination with consequential disturbances that push the parties to an interfirm agreement off of the contract curve. Strategic considerations—which arise by reason of information asymmetries, bilateral dependencies, weaknesses of property rights, and the costliness of court enforcement of contracts—will now come into play if, rather than mere frailty of motive, opportunism is the operative condition.

Adaptations.—Both the organization theorist Chester Barnard and the economist Friedrich Hayek took adaptation to be the main purpose of economic organization, but with differences. Finding little in the social sciences that informed the study of internal organization (hierarchy) as he had experienced it, Barnard undertook to craft the relevant concepts himself in his pathbreaking book, The Functions of the Executive (Barnard 1938), where he focused on coordinated adaptation as accomplished in a “conscious, deliberate, purposeful” way through the use of administration. Hayek, by contrast, celebrated the “marvel of the market” (Hayek 1945, 527) where autonomous adaptations are implemented spontaneously in response to changes in relative prices.

The challenge for the economics of governance was to recognize that adaptations of both kinds are important and to make selective provision for each. Rather, therefore, than be trapped in the old ideological divide between markets or hierarchies, transaction cost economics treats the two as alternative modes of governance, markets and hierarchies, both of which have distinctive roles to play in a well working economy. The heretofore maligned mode of hierarchy is now awarded coequal status with the marvel of the market, the object being to deploy each appropriately.

Contract Laws (Plural).—As against the standard practice of there being one all-purpose law of contract, Karl N. Llewellyn, who was a leader in the Legal Realism Movement in the United States, moved beyond the concept of contract as legal rules (with court enforcement) by introducing the idea of “contract as framework,” predominantly as implemented by private ordering. Specifically, the “major importance of legal contract is to provide … a framework which almost never accurately indicates real working relations, but which affords a rough indication around which such relations vary, an occasional guide in cases of doubt, and a norm of ultimate appeal when the relations cease in fact to work” (Llewellyn 1931, 736–737). This last condition is important, in that recourse to the courts for purposes of ultimate appeal serves to delimit threat positions. The more elastic concept of contract as framework nevertheless supports a (cooperative) exchange relation over a wider range of contractual disturbances than would a strict legal rules construction. As discussed below in conjunction with pushing the logic to completion, the contract law regime of “forbearance” has similar purposive origins.

Suffice it to observe here that adaptations (of autonomous and coordinated kinds) are taken to be the central purpose of organization; and viable modes of governance differ in contract law respects.

B. Key Operational Moves

The three key operational moves are to (i) name the attributes of the unit of analysis, (ii) do the same for modes of governance, and (iii) advance the efficient alignment hypothesis.
Various units of analysis have been proposed for studying organizations, yet efforts to name the defining attributes of proposed units are rare. The unit of analysis in the transaction cost economics set-up is the transaction—as recommended by Commons (1932) and as is implicit in Coase (1937, 1960). For transaction cost economizing purposes, the critical dimensions of transactions are complexity, the condition of asset specificity, and the disturbances to which a transaction is subject. As among these three, the attributes of transactions that have been most important to an understanding of the governance of contractual relations are the conditions of asset specificity and outlier disturbances for which unprogrammed adaptations are needed.10

Although Jacob Marschak had made perceptive reference to specialized human and locational conditions and observed that “the problem of unique or imperfectly standardized goods … had been neglected in the textbooks” (Marschak 1968, 14), the wide reach of asset specificity—to include physical, human, site specific, dedicated, brand name capital, and episodic (or temporal) forms—would become evident only as the concerted study of transaction cost economics got under way. Relevant in this connection is that different types of hazards accrue to different forms of asset specificity, which variations have significant organizational ramifications. Whatever the particulars, the basic regularity that is associated with transactions that are supported by investments in specific assets is that these assets cannot be redeployed to alternative uses and users without loss of productive value (Williamson 1971, 1975, 1976, 1985; Benjamin Klein, Robert G. Crawford, and Armen A. Alchian 1978).

Intertemporal considerations are relevant in this connection. Thus although some conditions of asset specificity are evident from the outset, others evolve during contract implementation. (Human asset specificity that arises because of learning by doing is an example of the latter.) Whatever the source of the condition of asset specificity, the condition of nonredeployability, to which I refer above, has the effect of transforming what may have been a large numbers bidding competition at the outset into a small numbers exchange relationship during the period of the contract and at the contract renewal interval. Such transformations compromise the efficacy of simple market exchange, which is supplanted by longer term contracts (as supported by credible commitments) or, in the limit, by unified ownership of successive stages with recourse to hierarchy.11

Markets and hierarchies are the two polar modes to which Coase referred in his 1937 paper and are the governance alternatives on which I focus in my paper on the vertical integration of production. This is wholly in the spirit of the first precept of pragmatic methodology: keep it simple (Robert M. Solow 2001; Brian Snowdon and Howard Vane 1997). It is noteworthy, however, that transaction cost economics has subsequently introduced the hybrid mode (Williamson 1991; Claude Menard 1996) and has furthermore moved beyond intermediate

10 Note, however, that complexity plays a crucial role in the following respect: all complex contracts are incomplete by reason of bounded rationality. Not all incompleteness, however, is consequential.

I associate consequential incompleteness mainly with outlier disturbances for which the stakes are great (because the parties are bilaterally dependent), on which account asset specificity and uncertainty are the defining features. Inconsequential incompleteness is that range of disturbances over which Llewellyn’s “contract as framework” can be presumed to work well, often with the support of credible contracting mechanisms.

It is also, however, useful to recognize that incompleteness becomes more severe as the number of features of a transaction (precision, linkages, compatibility) across which adaptations are needed increase and as the number of consequential disturbances that impinge on these features increase, which disturbances increase with the length of the contract.

11 Note that because asset specificity is a design variable, the good or service to be delivered could be redesigned so as to reduce asset specific features, albeit at a sacrifice in performance of the good or service in question (Michael H. Riordan and Williamson 1985). Note also that whereas the emphasis on individual (bilateral) transactions serves the purpose of analytical simplicity, groups of related transactions sometimes pose sequencing problems. This introduces systems considerations for which real time coordination complexities need to be factored in. (See note 18, infra.)
product market contracting to interpret a wide range of commercial (and some noncommercial) phenomena as variations on a theme.

The critical dimensions for describing alternative modes of governance, of which markets and hierarchies are two, are incentive intensity (which is strong in autonomous stages that appropriate their streams of net receipts and is weak under cost-plus reward schemes), administrative command and control (which is strong if successive stages are under unified ownership and are subject to coordination and dispute resolution by a common “boss”), and contract law regime, which is strong under a legal rules (court ordered) contract law regime but is weak if disputes between successive stages are settled by private ordering (where the firm is its own final court of ultimate appeal).

Assuming that each of these three dimensions of governance can take on either of these two values, weak (0) or strong (+), and that we focus on polar modes (market and hierarchy), there are \(2^3 = 8\) possible combinations. Which are the internally consistent combinations that describe market and hierarchy? As discussed elsewhere (Williamson 1991), the syndrome that describes the market is strong incentive intensity, weak command and control at the interface, and strong (legal rules) contracting. The syndrome that describes hierarchy, by contrast, is weak incentive intensity, strong administrative command and control at the interface, and weak contract law regime (forbearance law). So described, market and hierarchy are polar opposites.

Efficient Alignment.—Transaction cost economics appeals to the efficient alignment hypothesis to predict which transactions go where—to wit, transactions, which differ in their attributes, are aligned with governance structures, which differ in their cost and competences, so as to effect a (mainly) transaction cost economizing outcome. The basic prediction for generic transactions for which asset specificity is nil and the adaptive needs can be ascertained and implemented autonomously is that these will be procured in the market. Such transactions correspond to the ideal transactions in both law and economics. Transactions, by contrast, that are supported by significant investments in transaction specific assets and are subject to incompleteness (by reason of bounds on rationality) will experience malcoordination when beset by significant disturbances for which defection from cooperation can be projected as the stakes increase. Such transactions will benefit from unified ownership and coordinated adaptations as implemented by hierarchy.

C. Applications

Economic theories take on added significance if and as (i) the predictions are borne out by the data, (ii) variations on a theme are worked out, and (iii) public policy ramifications accrue and are displayed.

Empirical.—Transaction cost economics both makes predictions and submits them to empirical testing. Not only did empirical tests of transactions cost economics number over 800 in 2006, but they have been broadly corroborative (Jeffrey T. Macher and Barak D. Richman 2008). 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” (Inge Geyskens, Jan-Benedict Steenkamp, and Nirmalya Kumar 2006, 531). There is no gainsaying that transaction cost economics has been much more influential because of the empirical work that it has engendered (Michael D. Whinston 2001).

Variations on a Theme.—Transaction cost economics has many applications not only 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. Such broad reach arises because any problem that arises as or can be reformulated as a contracting problem can be examined to advantage in transaction cost economizing terms.

Public Policy. Although transaction cost economics has had numerous applications to public policy toward business (antitrust, regulation, corporate governance) and in some degree in the study of agriculture, public health, public bureaus, and economic development and reform, it is, in my judgment, an underused public policy perspective—especially in the design of public bureaus, of which the Department of Homeland Security in the United States is a recent example (Dara K. Cohen, Mariano-Florentino Cuellar, and Barry R. Weingast 2006). An efficiency assessment of feasible alternatives is too often scanted by a political process where public bureaus are designed with reference to immediate political purposes.

IV. Pushing the Logic to Completion

Pushing the logic to completion is accomplished by combining the second and third precepts of pragmatic methodology—namely, “get it right” and “make it plausible” (Solow 2001, 111). Getting it right “includes translating economic concepts into accurate mathematics (or diagrams, or words) and making sure that further logical operations are correctly performed and verified” (Solow 2001, 112); and plausible simple models of complex phenomena are expected to “make sense for ‘reasonable’ or ‘plausible’ values of the important parameters” (Solow 2001, 112). Also, because “not everything that is logically consistent is credulous” (David M. Kreps 1999, 125), fanciful constructions that lose contact with the phenomena are suspect—especially if alternative and more veridical models yield refutable implications that are congruent with the data. Combining precepts 2 and 3, the argument comes down to this: push the logic to completion, as tempered by considerations of feasibility.

Pushing the logic of zero transaction cost to completion with respect to externalities (Coase) and vertical integration (Arrow) revealed that routine recourse to this simplifying assumption led to counterfactual predictions, as a result of which economists and other social scientists were encouraged to push the logic of positive transaction costs to completion—both in general and as revealed by gaps or omissions that would become evident as the transaction cost economics setup evolved. Four such conditions are examined here: the impossibility of selective intervention, which bears on limits to firm size; the concept of remediableness, which has massive public policy ramifications by insisting on feasible, implementable solutions; credible contracting, which is a robust concept for expanding the range within which mutual gains from trade can be projected; and the test of scaling up to ascertain whether successive application of the simple (toy) model on which the analysis rests yields a scaled up version that approximates the phenomenon in question. Also, I briefly discuss the natural progression.

A. Selective Intervention

The limits to firm size puzzle, as posed by Frank Knight (1933) and Coase (1937), is this: Why can’t a large firm do everything that a collection of smaller firms can do and more? Tracy

Lewis answers a variant of this puzzle as follows: because an established firm can always “use the input exactly as the newer entrant would have used it …[and can furthermore] improve on this by coordinating production from his new and existing puts” the large firm will always realize greater value (Tracy R. Lewis 1983, 1092). Transaction cost economics examines this argument by postulating two mechanisms—replication and selective intervention—which, if they could be implemented, would support the all-purpose superiority of larger firms.

Thus suppose that two successive stages of production are combined with the understanding that (i) the acquired stage will operate in the same autonomous manner postacquisition as in the preacquisition status (by replication) except as (ii) the acquiring stage intervenes selectively, always but only when expected net gains can be ascribed to coordinated adaptations. In that event, the combined firm can never do worse (by replication) and will sometimes do better (by selective intervention). Accordingly, more integration is always better than less—which is to say that, upon repeated application of this logic, everything will be organized in one large firm.

Wherein does the implementation of this logic break down?

Assuming that the buyer stage acquires the supplier stage, the four conditions for implementing replication and selective intervention are these: (i) the buyer stage as acquirer (owner) promises the acquired supplier that the acquired stage will continue to appropriate its net receipts (as reduced by overhead, maintenance, user and capital depreciation charges) in all state realizations—thereby to preserve high powered incentives; (ii) the supplier promises to utilize the supply stage assets, the ownership of which has been transferred to the buyer, with “due care”; (iii) the buyer promises always and only to exercise authority (fiat) when expected net benefits can be ascribed to selective intervention; and (iv) the buyer promises candidly to reveal and divide the benefits that accrue to selective intervention as stipulated in the acquisition agreement. The problem is that none of these promises is self-enforcing. To the contrary, in the absence of three-way common knowledge (to include a costless arbiter) each condition will be compromised. Contributing factors include (i) 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, (ii) failures of due care become known only with delay and are difficult to prove, (iii) the buyer can also falsely declare state realizations to favor its own stream of net receipts, and (iv) in consideration of the foregoing, the division of benefits under selective intervention can be compromised. Also, (v) the political game is now played in a larger firm that is more susceptible to bureaucratic ploys and political positioning than in smaller firms.

The details of this brief sketch are reported elsewhere (Williamson 1985, Chap. 6). Suffice it to observe that the breakdowns referred to above are often intuited by many intelligent businessmen and their lawyers, who recognize the tradeoffs and factor them into the decision to integrate (or not). The lesson for social scientists is that markets and hierarchies differ in discrete structural ways, and we need to come to terms with the strengths and weaknesses of each.

B. Remediableness

The remediableness criterion serves as a reality check on the practice among public policy analysts of assuming that transaction costs in the public sector are zero. Not only is that nonsense, but 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

---

13 The need for three-way common knowledge, to include the arbiter, is yet another example of pushing the logic to completion (Williamson 1975, 21–34). The assumption that two-way common knowledge suffices is nonetheless widely held.
concept for public sector failures. 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, warts and all. The criterion is this: an extant 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).

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, non-profit, 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. The presumption that an extant mode is efficient if the expected net gain is negative can nevertheless be rebutted by showing that the obstacles to implementing an otherwise superior feasible alternative are “unfair.”

Fairness of both political and economic kinds come under review. Thus whereas political obstacles that are judged to be fair in circumstances where politics properly trumps economics (Stigler 1992) survive, those that have unacceptable political origins (e.g., are unfairly discriminatory) do not. Likewise, whereas some economic obstacles, such as sunk costs that have been incurred by the incumbent, may warrant delaying the introduction of a superior feasible alternative, those that are judged to be unfair (e.g., predatory behavior) will be challenged.

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

C. Credible Commitments

The concept of credible threat figures prominently in the study of rivalry (between nation states, in politics, and in business), where the main purpose of a credible threat is to deter the use of some instruments (e.g., nuclear weapons), thereby to deflect competition to other venues (Thomas C. Schelling 1960) or to deter the appearance of competition altogether. The use of cost effective credible commitments to support exchange is related but different.

The basic proposition is this: absent the use of credible commitments to support exchange, the contractual hazards associated with many transactions would be perceived to be excessive. Generic investments would replace transaction specific investments if the latter pose too great a risk. Some transactions would be taken into firms. Some would never materialize.

Credible commitments sometimes come into place spontaneously, as where a history of good experience with a trader leads to a positive reputation effect. Often, however, credible commitments take shape as economic actors consciously agree upon mechanisms that provide added assurance. These can take the form of information disclosure and auditing mechanisms, the development of specialized dispute settling mechanisms, whereby the parties rely more on private ordering than court ordering (Llewellyn 1931; Stewart Macaulay 1963; Clyde W. Summers

---

14 Albeit a caricature, “normative public policy analysis began by supposing that… policy was made by an omnipotent, omniscient, and benevolent dictator” (Avinash K. Dixit 1996, 8)—which, in transaction cost terms, assumes the absence of implementation obstacles, bounds on rationality and opportunism, respectively.

15 To be sure, unfair obstacles to implementation may persist even after a showing that these stand in the way of progress. Obstacles to efficiency nevertheless invite dissent. Some can be overturned by the cumulative force of movements, of which the civil rights movement is an example, and others by perfecting definitions of unfair competition.

16 The 32-year coal supply agreement between the Nevada Power Company and the Northwest Trading Company is illustrative (Williamson 1991).
Credibility supports also vary with the institutional environment as among political jurisdictions (Brian Levy and Pablo T. Spiller 1994), to which the literature on positive political theory is relevant. Also relevant to the economics of governance is the concept of contract laws (plural), an example of which is the concept of “forbearance law” to describe the contract law regime within hierarchy (Williamson 1991, 274; emphasis added):

The implicit contract law of internal organization is that of forbearance. Thus, whereas courts routinely grant standing to firms should there be disputes over prices, the damages to be ascribed to delays, failures of quality, and the like, courts will refuse to hear disputes between one internal division and another over identical technical issues. Access to the courts being denied, the parties must resolve their differences internally. Accordingly, hierarchy is its own court of ultimate appeal.

The concept of forbearance law regime was introduced to fill a logical gap in the theory of governance. As with other forms of contract law, the efficacy of forbearance law varies with the integrity of the institutional environment (nation state) of which it is a part.

D. Scaling Up

The object of a simple model is to capture the essence, thereby to explain hitherto puzzling practices and make predictions that are subjected to empirical testing. Often, however, 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 usually ignored, possibly out of awareness that scaling up cannot be done. Sometimes it is scant, possibly in the mistaken belief that scaling up can be accomplished easily. My position is that claims of real world relevance, including public policy relevance, of any proposed theory of the firm that cannot be shown to scale up from toy model status to approximate the phenomenon of interest (e.g., the modern corporation) should be regarded with caution.

With respect to the transaction cost 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 transaction cost economics set-up, scale up to describe something that approximates a multistage firm? Note that transaction cost economics assumes that the transactions of principal interest are those that take place at the interface between (rather than within) technologically separable stages. Upon taking the technological “core” as given, one focuses attention

---

17 Efforts to enhance credibility sometimes take strange forms, presumably because the parties are unable to do better. Thus consider the recently unearthed tablets in Mesopotamia (dated around 1750 B.C.) which reveal that self-inflicted curses were used to deter the breach of treaties. One of these reads as follows:

When you ask us for troops, we will not withhold our best forces, we will not answer you with evasions, we shall brandish our maces and strike down your enemy …

As wasted seeds do not sprout, may my seed never rise, may someone else marry my wife under my eyes, and may someone else rule my country (China Daily, March 22, 1988, p. 1).

18 Note that forbearance law precludes court jurisdiction over most internal decision making to which internal consequences accrue, but court jurisdiction does apply to externalities.

19 Michael C. Jensen and William H. Meckling (1976) posed the question of whether their simple model of entrepreneurial ownership scaled up to deal with the diversely owned modern corporation. They conjectured that it did apply but deferred a demonstration of scaling up to a later paper. Alas, that paper never appeared. Jensen and Meckling nevertheless candidly perceived the need for scaling up.
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).\footnote{There is, however, a caveat: scaling up, so described, does not make allowance for systems complications of the kind that arose in conjunction with Boeing Aircraft’s production of the 787 Dreamliner, where outsourcing confusion was rampant (Sanders, Peter 2009. “Boeing CEO’s Bumpy Ride.” Wall Street Journal, Nov. 5. http://online.wsj.com.). With the benefit of hindsight, malcoordination among outsourced transactions led to costly delays which possibly could have been averted if related components for which real time coordination would prove to be crucial had been produced internally. The requisite apparatus to address the systems complications that can arise among groups of related transactions has yet to be worked up within transaction cost economics. Applications of transaction cost economics would, however, have avoided the most serious outsourcing error made by Boeing: the decision to outsource the highly specialized fuselage to Vought Aircraft Industries. This transaction required significant investments in specific assets and would pose a series of adaptation problems during contract implementation (Tadelis 2010a). Boeing subsequently rectified this condition by acquiring Vought (Sanders, 2009. “Boeing Takes Control of Plant.” Wall Street Journal. December 23, p. B2).}

E. The Natural Progression

Transaction cost economics is sometimes criticized because it has not been fully formalized, to which I have three responses: transaction cost economics, like many other theories, has undergone a natural progression; full formalization is a work in progress; and premature formalization runs the risk of a disconnection with the phenomena.

Theories commonly progress from informal to preformal to semiformal to fully formal stages of development—broadly in the spirit of Thomas S. Kuhn (1970). The informal stage of transaction cost economics was the literature from the 1930s (especially Commons and Coase) where errors or omissions in the neoclassical set-up were described. Preformal work got under way in the 1970s, where new concepts for reinterpreting vertical integration, vertical market restrictions, labor market organization, franchise bidding for natural monopoly, and the like were forged and the conditions for efficient alignment were worked out. Semiformal work, in the 1980s and since, deal with credible contracting, hybrid modes, the dimensionalization of transactions and governance structures, a multiplicity of applications within business and economics and the contiguous social sciences (to include public policy), and extensive empirical testing. Full formalism also got under way in the 1980s and is still in progress. The pathbreaking paper by Sanford J. Grossman and Oliver D. Hart (1986) and the follow-on paper by Hart and John Moore (1990) and others in this tradition—which deal with some types of transaction costs (but is more often referred to as the property rights literature)—have been very influential. Subsequent significant work by Steven Tadelis and his coauthors (Patrick Bajari and Tadelis 2001; Tadelis 2002; Levin and Tadelis forthcoming; Tadelis 2010a) is likewise in progress.

V. Concluding Remarks

What I describe as the transaction cost economics project had its origins in the puzzle posed by Coase in 1937: What explains the boundaries of the firm? I addressed this by taking the vertical integration decision to be the focal transaction and, upon reformulating it as a contracting problem, asked the following question: When and why should a firm acquire a technologically separable component by outsourcing rather than producing to its own needs—where outsourcing entails contracting out and own-production to contracting within. This question was addressed as an efficiency issue by selectively combining economics with organization theory. Albeit intended
as a standalone research project, the vertical integration set-up would open windows to a wide array of economic activities that arose as or could be reformulated in comparative contractual terms.

With the benefit of hindsight, transaction cost economics has undergone a natural progression. The informal stage got started in the 1930s with Coase’s challenge to the profession that firm and market organization should be derived rather than (as was then the practice) taken as given, to include the suggestion that the missing concept was transaction cost. This latter was buttressed by demonstrations (by Arrow and Coase) in the 1960s that much of standard economics was reduced to irrelevance upon pushing the logic of zero transaction costs to completion.

The preformal stage began in the 1970s with the application of the lens of contract/governance to vertical integration. Interfirm contracts that were incomplete (by reason of bounds on rationality) would experience maladaptation hazards if the parties were bilaterally dependent (by reason of transaction specific investments) in the face of disturbances for which the stakes are great (strategic defection). This economics of governance approach would subsequently have wide application as other contractual phenomena were interpreted as variations on a theme.

The semiformal stage gave added prominence to the defining attributes of alternative modes of governance (market, hybrid, and hierarchy) as these relate to differing adaptive needs, of autonomous and coordinated kinds, among different transactions. A series of puzzles arose as this operationalization effort progressed—of which the efficacy of selective intervention was one and scaling up was another, for which the answers would be revealed by pushing the logic of economic organizations to completion. Beginning in the early 1980s and growing exponentially thereafter, an ambitious effort at empirical testing got under way. Applications to public policy are likewise numerous and growing. Fully formal research of a transaction cost economics kind has taken shape, and more is in progress.

I conclude that selectively combining law, economics, and organization to study the governance of contractual relations from a transaction cost economizing perspective has been instructive; and I project that research of this kind will continue to develop in conceptual, theoretical, empirical, and public policy respects. Research in transaction cost economics faces an interesting, challenging future.

REFERENCES


