

## The Economics of Governance

Oliver E. Williamson

<owilliam@haas.Berkeley.edu>

University of California, Berkeley

September 27, 2004

This paper is a draft and should not be quoted without the permission of the author.

## The Economics of Governance

Oliver E. Williamson\*

Governance, as herein described, is the means by which to infuse order, thereby to mitigate conflict and realize mutual gains from trade. The most obvious candidate, at least among economists, for achieving these gains is that of simple market exchange, according to which “faceless buyers and sellers ... meet ... for an instant to exchange standardized goods at equilibrium prices” (Ben-Porath, 1980, p. 4). But while that is an obvious place to begin, the economics of governance is interested not merely in simple but also in complex transactions, to include those that are organized in nonmarket as well as market modes of governance. To the “marvel of the market” to which Friedrich Hayek (1945) made perceptive reference is therefore added the “marvel of hierarchy” (of both private and public sector kinds). The overarching argument is this: alternative modes of governance, which possess different strengths and weaknesses, are aligned with the adaptive needs of different transactions --the upshot of which is that there is a place for each generic mode of governance, yet each should be kept in its place.

As discussed elsewhere (Williamson, 1998, pp. 598-599), governance is usefully examined in relation to two other levels of institutional analysis: the informal and formal rules of the game – namely, the societal level (norms, customs, and the like) and the institutional environment (the polity, laws of property and contract, and their enforcement), respectively.

Feedback effects aside, governance is the play of the game as constrained by these informal and formal rules of the game.

So construed, the economics of governance works off of a series of promising ideas that were advanced, mainly independently, in the 1930s – by Karl Llewellyn (1931), John R. Commons (1932), Ronald Coase (1937), and Chester Barnard (1938). Partly because the ideas were so novel and the connections between them were not perceived, the operationalization of these good ideas was slow to materialize. The economics of governance is an effort to make the requisite connections and breathe operational content into the project.

As herein described, the economics of governance adopts a lens of contract perspective and focuses on the microanalytics of transactions on one hand and of alternative modes of governance (markets, hybrids, hierarchies, bureaus, etc.) on the other. Complex contracts are assumed to be incomplete and adaptation to disturbances is taken to be the central problem of economic organization. The resulting approach to economic organization differs from both “contract economics” (agency theory, mechanism design, and formal property rights theory)<sup>1</sup> and orthodoxy (the resource allocation paradigm). The contrasts with orthodoxy are manifold: governance examines economic phenomena at a more microanalytic level of detail than does orthodoxy, focuses on positive transaction costs (rather than assume these to be zero), describes the firm as a governance structure (an organizational construction) rather than a production function (a technological construction), is more interdisciplinary, in that law and organization theory are selectively combined with economics, and works out of

an efficiency rather than a monopoly presumption for interpreting nonstandard and unfamiliar contracting practices and organizational structures.

The upshot is that the economics of governance employs a different gestalt, for which, of course, there are setup costs (Kuhn, p. 204).<sup>2</sup> The prospective benefits reside in deeper understandings, additional refutable implications, and better public policy ramifications.

The core concepts on which the economics of governance are based are set out in Section 1. Additional conceptual moves that are needed to “think contractually” about economic organization are described in Section 2. The pragmatic methodology out of which the economics of governance works is sketched in Section 3. The operationalization of governance is accomplished by focusing on the microanalytics of contract and organization in the context of vertical integration, which has become the paradigm problem, in Section 4. Variations on a theme, public policy toward business, and empirical testing are examined in Section 5. Tensions and challenges are addressed in Section 6. Concluding remarks follow. As will be apparent throughout, the economics of governance is a work-in-progress.

#### 1. Core Concepts I: The Foundations

The three core concepts discussed here – the lens of contract/governance, human actors, and transaction cost – all have their origins in the 1930s. Because, however, these concepts were introduced a very high level of generality, the ramifications for economic organization remained obscure.

I attempt to add economic content by (1) moving to a higher level of resolution and (2) yoking the complementary concepts of governance and transaction cost, which in combination serve to complete each other.<sup>3</sup>

### 1.1 the lens of contract/governance

James Buchanan (1975) contrasts the orthodox lens of choice (the resource allocation paradigm) with the much less fully developed lens of contract<sup>4</sup> and avers that “mutuality of advantage from voluntary exchange ... is the most fundamental of all understandings in economics” (2001, p. 29). The economics of governance is a lens of contract construction, with emphasis on the ex post governance of ongoing contractual relations.

John R. Commons was one of the leaders of the older style of institutional economics in the U.S. By contrast with orthodoxy, which aspired to work out of an “institution free core,” institutional economists held that institutions mattered. Because, however, it takes a theory to beat a theory and because a positive research agenda was not forthcoming (Stigler, 1983), that project “ran itself into the sand.”<sup>5</sup>

Such a fate notwithstanding, institutional economics was onto good issues. Of these, none was more important in my judgment than Commons’ abiding interest in “going concerns.” As against the preoccupation of orthodoxy with simple market exchange and the resource allocation paradigm (Reder, 1999), Commons observed that the continuity of an exchange relationship was often important, whereupon the problem of economic organization was

reformulated 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, p. 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” (1950, p. 21).

Although a coherent theory of organization for implementing these novel concepts eluded Commons and his followers -- possibly because the concept of transaction cost had yet to surface and because of the primitive state of organization theory at the time<sup>6</sup> --, the Commons triple of conflict, mutuality, and order anchors the concept of governance as herein employed. Specifically, governance is the means by which to infuse order, thereby to mitigate conflict and realize the mutual gains from trade to which Buchanan makes reference.

## 1.2 human actors

Herbert Simon advised social scientists that “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” (1985, p. 303). The two attributes of human actors that are especially relevant to the economics of governance are cognition and self-interestedness.

Simon took early exception with hyperrationality and proposed instead that human actors be described as boundedly rational, by which he means that they

are “intendedly rational, but only limitedly so” (1957a, p. xxiv). Human actors are thus neither nonrational nor irrational but are attempting effectively to cope.

Gerd Gigerenzer has since split bounded rationality into two parts: “satisficing”, which refers to searching for an alternative that is “good enough” (Simon, 1957b, p. 202), and “fast and frugal heuristics,” such as routines. My own take on bounded rationality is that it is an encompassing concept that “cannot be precisely defined. It is a problem that needs to be explored” (Selten, 2001, p. 15) – to which I would add that such explorations should take place in the context of specific phenomena. With reference to complex board games (such as chess), bounds on rationality are relieved by heuristics; in the context of search it is manifested as satisficing; my interest is in contracting, where the main lesson of bounded rationality is that all complex contracts are unavoidably incomplete.

Comprehensive contingent claims contracting is thereby disallowed. Adaptive, sequential decision making arises instead, which is to say that all possible bridges are not crossed in advance but many are crossed if and as they arise. This does not, however, imply that successive contingencies are dealt with in an ad hoc way. Rather, the economics of governance makes the additional cognitive assumption that parties to a long-term contract possess “feasible foresight,” by which I mean that they have the capacity to look ahead, uncover possible hazards, and work out the ramifications, thereupon to incorporate hazard mitigating mechanisms within the ex ante contractual agreement. Not

only are there distinguished antecedents for such an assumption – both in sociology (Michels, 1962, p. 370) and economics (Shultz, 1995, p. 1; Arrow, 2000, p. 12) –, but the predictions of credible contracting and long-term contracting in practice are corroborative. Were it that more practitioners in the contiguous social sciences were persuaded to replace myopia with feasible foresight, at least in the context of commercial contracting, many of the existing tensions between economics and the contiguous social sciences would be relieved.

### 1.3 transaction cost

The orthodox position on transaction costs is that these are negligible and can be assumed to be zero (Coase, 1960, 1992; Arrow, 1969). In that event, the costs of contracting and organization are ignored and attention is focused instead on prices and quantities, supply and demand. The resulting resource allocation paradigm became the “dominant paradigm” within economics for the 20th century (Reeder, 1999).

It is uncontested that Ronald Coase had a talent for uncovering lapses in what others regarded as unproblematic.<sup>7</sup> His path-breaking paper on “The Nature of the Firm” (1937) is the first and arguably the most important of his challenges to orthodoxy. Coase’s training in business administration and his field work on vertical integration contributed to his skepticism with orthodox treatments of firm and market organization (Coase, 1988, pp. 5-8). He therefore resolved to address the puzzle, “What was it that determined whether [to supply a good or

service] by contractual arrangements between independent firms or by integration?” (Coase, 1988, pp. 16-17). The solution upon which he hit “was to realize that there were costs of making transactions in the market and that it was necessary to incorporate them into the analysis. This was not done in economics at that time – nor, I may add, is it in most present-day economic theory” (Coase, 1988, pp. 16-17).

Coase’s 1937 article confronted orthodox economics with a logical lapse. Thus whereas orthodoxy took the distribution of economic activity as between firm and market organization as given, whereupon attention was focused on “the economic system as being coordinated by the price system” (1937, p. 387), firm and market are properly regarded as “alternative methods of coordinating production” (1937, p. 388; emphasis added). Rather than take the distribution of economic activity as given, this should be derived. The 1937 paper thus took as its purpose “to bridge what appears to be a gap in economic theory.... We have to explain the basis on which, in practice, this choice between alternatives is effected” (Coase, 1937, p. 389, emphasis added).

These arguments posed a major challenge to economics over the next 35 years, yet failed to take hold (Coase, 1972, p. 63). A major obstacle is that vaguely described transaction costs came to be invoked as an all-purpose explanation for puzzling contractual and organizational practices, as a result of which transaction cost reasoning acquired a “well-deserved bad name” (Fischer, 1977, p. 322, n. 5).

Bad name notwithstanding, the concept of transaction costs plainly resonated with the experience of many economists and was too fundamental to let go. Might the tautological reputation of transaction cost reasoning be overturned by focusing on contractual specifics? That is what I attempted to do by re-examining vertical integration “on its own terms” (Williamson, 1971).<sup>8</sup>

To be sure, this took place in the context of recent developments in the market failure literature (Coase, 1960; Arrow, 1969), which encouraged me to believe that a comparative contractual lens is an instructive way by which to examine puzzling economic phenomena. Relatedly, an efficiency orientation was adopted: if one form of organization is used rather than another, entertain the possibility that efficiency gains are associated with the winner. I furthermore made allowance for the limits of human actors in cognitive and trustworthiness respects. Going beyond simple market exchange, the final move was to examine whether intertemporal process transformations are (sometimes) associated with ongoing contractual relations. Taking the transaction to be the basic unit of analysis and naming adaptation (of both autonomous and cooperative kinds) as the main purpose of economic organization served to focus attention on an important but hitherto neglected class of transaction costs: maladaptation.

Like production costs, transaction costs can take many forms: search costs, measurement costs, administrative costs, information processing costs, maladaptation costs. The existence of any such costs is not, however, where the analytical action resides. Whether a class of transaction costs has ramifications for the boundary of the firm turns on answers to two questions. First, do

alternative modes of governance differ in their capacities to cope with such costs? Second, what are the attributes of transactions that are responsible for variations in such costs? Differences among alternative modes of governance of a discrete structural kind explain capacity differences. The complexity of transactions, somehow described, presumably explain cost variations.

Upon combining the Commons propositions that the transaction is the basic unit of analysis and that governance is the means by which to infuse order, thereby to mitigate conflict and realize mutual gains with the Coasian insight that transaction cost economizing is the missing criterion for judging efficacy, operational life is breathed into both.<sup>9</sup>

## 2. Thinking Contractually: Core Concepts II

Upon adopting a comparative contractual approach to the study of economic organization and naming the transaction as the basic unit of analysis, it also becomes natural to examine the challenges posed by different transactions to which governance is a response. This section examines these issues.

I observe in this connection that it is not by design but by the nature of the problem that the economics of governance moves beyond economics to selectively draw upon law and organization theory. The question is this: if and as problems can be contained within economics only by making heroic assumptions, can selective appeal to the contiguous social sciences restore

relevance? Many economists are understandably reluctant to make such a move (Lazear, 2000, pp. 99-100, 103).<sup>10</sup>

## 2.1 adaptation

Interestingly, both the economist Friedrich Hayek (1945) and the organization theorist Chester Barnard (1938) were in agreement that adaptation is the central problem of economic organization. Hayek (1945, pp. 526-527) focused on the adaptations of economic actors who adjust spontaneously to changes in the market, mainly as signaled by changes in relative prices: Upon looking “at the price system as...a mechanism for communicating information,” the marvel of the market resides in “how little the individual participants need to know to be able to take the right action.” By contrast, Barnard featured coordinated adaptation among economic actors working through administration (hierarchy). The latter is accomplished not spontaneously but in a “conscious, deliberate, purposeful” way (1938, p. 9) and comes into play when the simple market exchanges on which Hayek focused break down.<sup>11</sup>

In effect, the adaptations to which Hayek refers are autonomous adaptations in which individual parties respond to market opportunities as signaled by changes in relative prices whereas the adaptations of concern to Barnard are cooperative adaptations accomplished through administration within the firm. Because a high performance economic system will display adaptive capacities of both kinds, an understanding and appreciation for both markets and hierarchies (rather than the comparative economic systems dichotomy between

markets or hierarchies) is needed. The firm for these purposes is described not as a production function (which is a technological construction) but as a governance structure (which is an organizational construction). And the market is described similarly. The lens of contract, as against the lens of choice, is made the cutting edge.

One of the advantages of focusing on adaptation as the main case is that it brings added meaning to the idea of mutual gain. It is elementary that gains from trade will always be realized by moving onto the contract curve. But how is this to be accomplished in a world where complex contracts are incomplete and are implemented over time in the face of disturbances for which contingent provisions either have not been made or, if made, are often in error? Crafting governance structures that are attuned to the hazards and help the parties to restore efficiency (return to the shifting contract curve), where otherwise a costly impasse would develop, is needed in these circumstances. More attention to designing processes that have good adaptive properties (and less to concentrating all of the action in the ex ante incentive alignment stage) is thus one of the central lessons of the economics of governance.

## 2.2 intertemporal transformations

As Philip Selznick observed, organizations, like the law, “have a life of their own” (1945, p.10). The idea that organizations predictably undergo intertemporal transformations has ramifications for an understanding of both governance and the origins of transaction costs.

One of the earliest and most perceptive illustrations of an intertemporal transformation is Robert Michels's famous study of Political Parties ([1911], 1962), where he advanced the Iron Law of Oligarchy: "It is organization which gives birth to the dominion of the elected over the electors, of the mandatories over the mandators, of the delegates over the delegators. Who says organization, says oligarchy" (1962, p. 365). Interestingly, Michels uncovered the propensity for oligarchy to materialize while studying democracy. Original democratic intentions notwithstanding, the initial leadership develops attachments for office and is strategically situated in relation to the membership. The leadership thus entrenches itself by controlling information, manipulating rewards and punishments, and mobilizing resources to defeat rivals. Even worse, the entrenched leadership uses the organization to promote its own agenda at the expense of the membership.

One response to unwanted oligarchical outcomes would be to eschew all proposals to introduce democracy in favor of the status quo, but that is extreme. A more constructive and deeper lesson is to make allowance for oligarchy at the outset. Upon looking ahead and recognizing the propensity for oligarchy, it may be possible to mitigate foreseeable oligarchical propensities at the initial design stage. That is Michels's position: only a serene and frank understanding of the propensity for oligarchy will permit the student of politics to take informed actions to mitigate the hazards (1962, p. 370). More generally, the lesson is this: students of economic organization need to be advised of all significant intertemporal transformations whatsoever – thereafter to mitigate unwanted

consequences and augment the latent benefits. As heretofore described, this entails the exercise of feasible foresight.

The two most consequential intertemporal transformations that are uncovered by the economics of governance are the Fundamental Transformation and bureaucratization.<sup>12</sup> Both show up upon examining incomplete contracts in their entirety.

The former discloses that outsourcing can “pose problems ... if either (1) efficient supply requires investment in special-purpose, long-life equipment, or (2) the winner of the original contract acquires a cost advantage, say by reason of first mover advantages (such as unique location or learning, including the acquisition of undisclosed or proprietary technical and managerial procedures and task-specific labor skills)” (Williamson, 1971, p. 116). Such problems are uncovered by examining the contracting process in an intertemporal way and by recognizing that asset specificity (which can take a variety of forms) gives rise to bilateral dependency.

Such dependency arises by reason of the Fundamental Transformation, whereby a large numbers supply condition at the outset is transformed into a small numbers exchange relation during contract implementation and at the contract renewal interval. Because transactions for which bilateral dependency is in prospect pose potential conflicts when incomplete contracts are subject to unanticipated disturbances, added governance supports are introduced, in cost effective degree, for such transactions. The upshot is that whereas vertical integration that lacked a “physical or technical” basis had previously been viewed

as deeply problematic (Bain, 1968, p. 381), the possibility that vertical integration serves efficiency purposes (of a contractual rather than technical kind), through the added governance supports so provided, must now be entertained.

Also, the economics of governance pushes beyond organization theory in addressing the intertemporal burdens of bureaucracy that beset internal organization. There are three steps to the argument. First, although organization theory reports that numerous costly practices attend administrative forms of organization, this does not, without more, imply that internal organization is at a disadvantage to the market for these reasons. That is because trade always takes place between firms, all of which are hierarchically organized. The question, therefore, is not the existence of bureaucratic practices but whether the acquisition of one firm by another (say of a supplier by a buyer) leads to an increase in the aggregate bureaucratic costs of the combined entity. Second, and relatedly, the economics of governance addresses the question of whether a large firm can always do better than a collection of small firms, which is a widely held view. (This is addressed below in conjunction with the impossibility of implementing replication and selective intervention.)

The third step is to examine the differential consequences for politicizing economic organization that accrue upon taking a contract out of the market and organizing it internally. Integration implies “playing the game” of procurement within a larger administrative arena, which adds to the costs of bureaucratization in two respects: career concerns (which foster “going along”)<sup>13</sup> and sub-goal pursuit (as manifested by logrolling) (Williamson, 1985, Chap. 6).

### 2.3 contract laws (plural)

The ideal transactions, for both law and economics, are the simple market exchanges to which I referred at the outset. All well and good, for both law and economics, to begin the study of exchange by describing homogeneous transactions with large numbers of informed buyers and sellers on each side of the market. But that is merely a beginning. Not all transactions can be so described.

Going beyond simple market exchange, neoclassical economics focused on price theoretic complications due to monopoly, monopolistic competition, and externalities. The economics of governance calls attention instead to transactions that pose an altogether different class of complications: those for which the preservation of an ongoing contractual relation is the source of potential gains. Whether such gains will be realized or will be compromised by failures of cooperation and coordination becomes the issue.

To be sure, the fiction that contracts are well defined and costlessly enforced by well-informed courts is an analytical convenience for both law and economics. This normative ideal was nevertheless disputed by Karl Llewellyn, who held that a legal rules conception of contract was too narrow and introduced the idea of “contract as framework.” As Llewellyn put it, the “major importance of legal contract is to provide ... a framework which never accurately reflects real working relations, but which provides 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” (1931, pp. 736-737). The object of contract, so construed, was not to be legalistic but to get the job done.

The norm of ultimate appeal to which Llewellyn refers is important, in that recourse to the courts for purposes of ultimate appeal serves to delimit threat positions. But the key idea is this: the legalistic view of contract that applies to simple transactions needs to make way for a more flexible and managerial conception of contract as contractual complexities build up. As against the convenient notion of one all-purpose law of contract (singular), the need for contract laws (plural) is introduced. Such contract law differences play an important role in distinguishing among alternative modes of governance.

Marc Galanter’s contrast between “legal centralism” and “private ordering” is pertinent. Thus whereas the former maintains that “disputes require ‘access’ to a forum external to the original social setting of the dispute [and that] remedies will be provided as prescribed in some body of authoritative learning and dispensed by experts who operate under the auspices of the state” (Galanter, 1981, p. 1), the facts disclose otherwise. Most disputes, including many that under current rules could be brought to a court, are resolved by avoidance, self-help, and the like (Galanter, 1981, p. 2).

The unreality of the assumptions of legal centralism can, of course, be defended by reference to the fruitfulness of the pure exchange model. That is not disputed here. My concern is that the law and economics of private orderings are pushed into the background in the process. That is unfortunate, since in “many instances the participants can devise more satisfactory solutions to their

disputes than can professionals constrained to apply general rules on the basis of limited knowledge of the dispute” (Galanter, 1981, p. 4). The analytical convenience of black letter law and its costless enforcement thus gives way to a more purposive view of contract, with emphasis on cooperative adaptation, as complexities set in (Llewellyn, 1931; Macauley, 1963, Summers, 1969; Macneil, 1974).

The economics of governance resonates with the concept of contract laws (plural) and avers that each generic mode of governance is supported by a distinctive form of contract law. The contract law of simple market exchange is that of legal rules. The hybrid mode of contracting is supported by contract as framework. But what is the contract law of internal organization?

The argument here is that the implicit contract law of internal organization is that of forbearance. Thus, whereas courts routinely grant standing to contracts between firms should there be disputes over prices, the damages to be ascribed to delays, failures of quality, and the like, the courts have the good sense to 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, which is to say that the firm becomes its own court of ultimate appeal. The effect of forbearance law is to validate hierarchy by supporting its main purpose, namely timely responsiveness to disturbances for which coordinated adaptations are needed.

#### 2.4 three-way common knowledge

One way of formalizing the study of complex contracting and organization within the confines of economics is to assume that the parties to a transaction have common knowledge of payoffs and costlessly bargain to an efficient result in the ex post contract realization interval (Hart, 1995). In that event, all problems of inefficiency during contract implementation are vaporized, the analytical action having been concentrated instead in the ex ante incentive alignment stage of contracting. This gambit nevertheless encounters two obstacles. First, ex post common knowledge of payoffs is deeply problematic (Kreps and Wilson, 1982). Second, and more in the spirit of institutional analysis, two-way common knowledge does not yield ex post efficiency if the principals can appeal disputes to an arbiter to whom they can make false and misleading state realization reports. If it is costly for the arbiter to evaluate rival representations, then the assumption of two-way common knowledge must be extended to three-way common knowledge, to include the arbiter, if inefficiency in the ex post contract implementation stage is to be annihilated (Williamson, 1975, pp. 31-35). Another way of putting it is that two-way common knowledge is salvaged only if costless verifiability is ascribed to the arbiter, which is heroic compounded.

To be sure, one of the merits of formalization is that the need for strong assumptions is better disclosed. But that is merely the first step. Because a model can be right mechanically yet be “imperfectly suited to the subject matter” because key interactions are obscured (Solow, 2001, p. 112), the plausibility of

the assumptions should be addressed. Consistency is a weak test, in that “not everything that is logically consistent is credulous” (Kreps, 1999, p. 125).

## 2.5 discrete structural analysis

A chronic unanswered puzzle for the economics of organization is what is responsible for the limits to firm size. Tracy Lewis’s remarks on the purported coordination advantages of large established firms are pertinent (1983, p. 1092; emphasis added):

... [a] leader can at least use [an] input exactly as the entrant would have used it, and earn the same profits as the entrant. But typically, the leader can improve on this by coordinating production from these new and existing inputs. Hence the new input will be valued more by the dominant firm.

Expressed in terms of vertical integration, if a combined firm can always do better than a collection of small firms (by replication) and can sometimes do better (by coordination), then the limits to firm size must reside in unexamined obstacles to replication and selective intervention.

Thus suppose that firm A proposes to acquire firm B, which is a supplier to firm A, on the following terms: (1) firm A offers to pay M dollars to firm B for its physical assets; (2) firm A promises that firm B will appropriate its net receipts, which are defined in the post-acquisition interval as the transfer price times the quantity delivered less user costs, depreciation charges, and overhead; (3) firm B is directed to continue to operate just as it had in the pre-merger interval

(replication) except for (4) state realizations that push the parties off of the contract curve, in which event firm B will accede to selective intervention by firm A whenever prospective net gains to coordination can be projected. Examining the microanalytics of implementation, so described, quickly discloses that such a merger is beset with a series of contractual hazards.

For one thing, stage A depends on stage B to utilize and maintain assets (which it no longer owns) with the same degree of care in the post-merger interval and furthermore to make both process and product innovations in an unchanged way. Stage B has even greater concerns. It depends on stage A to make the appropriate adjustments in the transfer prices as conditions change, be meticulously fair in adjusting user costs, depreciation, and overhead charges, never to terminate or penalize the management in stage B except for due cause, and never to intervene selectively to require coordinated adaptations except as these are based on objective (non-strategic) reports of state realizations. As is obvious, and as elaborated elsewhere (Williamson, 1985, Chap. 6), this is an impossible list of conditions.

Because of the implausibility of both replication and selective intervention, those who aspire to study comparative economic organization in an uncontrived way must face the need to uncover the discrete structural differences that distinguish markets and hierarchies and work out the performance ramifications. As described below, the economics of governance holds that each generic mode of governance is described by an internally consistent syndrome of attributes – incentive intensity, administrative controls, and contract law regime –, as a

consequence of which each is well-suited to manage some transactions but poorly suited for others.

### 3. Pragmatic Methodology

The four precepts of pragmatic methodology are to keep it simple, get it right, make it plausible, and derive refutable implications to which the relevant microanalytic data are thereafter applied. Keeping it simple is made necessary by “the very complexity of real life” (Solow, 2001, p. 111), whence attention is focused on first order effects – the “main case,” as it were – after which qualifications, refinements, and extensions can be introduced. Getting it right entails working out the logic – to include “translating economic concepts into accurate mathematics (or diagrams, or words)” (Solow, 2001, p. 112). Making it plausible entails describing human actors in (reasonably) veridical ways, which serves as a check upon fanciful constructions.

The fourth precept of pragmatic methodology is to derive refutable implications to which the relevant (often microanalytic) data are brought to bear. To be sure, “the purpose of science in general is not prediction, but knowledge for its own sake” (Georgescu-Roegen, 1971, p. 37), yet prediction is the honest broker: it is “the touchstone of scientific knowledge” (Georgescu-Roegen, 1971, p. 37).

Some scoff at prediction, evidently in the belief that prediction is easy. Also, since everyone knows that “it is easy to lie with statistics,” what useful purpose is served by empirical testing? My experience is different: prediction is

a demanding standard, which is why so many would-be theories remain “excogitated speculations;” and corroboration is difficult, which explains why few predictions are tested.

Why more social scientists are not insistent upon deriving refutable implications and submitting these to empirical tests is a puzzle. One possibility is that the world of pure theory has its own orbit and rules of the game (Lipsey, 2001). A second possibility is that some theories are truly fanciful and their protagonists would be unmasked by such disclosure. A third is that the refutable implications of some would-be theories are contradicted by the data. A multiplicity of theories, some of which are vacuous, others of which are fanciful, and still others of which are contradicted by the evidence is an embarrassment to pragmatically oriented social scientists. Among this subset, insistence upon the fourth injunction – derive refutable implications and submit these to the data – has attractions.<sup>14</sup>

Progress, moreover, is realized by “weeding out empirically unsuccessful models and improving and extending those that survive empirical tests” (Solow, 1997, p. 50), in which event the theory and the evidence are often interactive, which is also Alan Newell’s position (1990, p. 14):

Theories cumulate. They are refined and reformulated, corrected and expanded. Thus, we are not living in the world of Popper.... [Theories are not] shot down with a falsification bullet.... Theories are more like graduate students – once admitted you try hard to avoid flunking them out.... Theories are things to be nurtured and changed and built up.

The idea that theories are progressively developed suggests the possibility of a natural progression – from informal, to pre-formal, semi-formal, and, finally, to fully formal modes of analysis. The informal stage is where the initial challenge to orthodoxy and early intuitions reside, the logic for which is thereafter developed in words (pre-formal), diagrams (semi-formal), and mathematics (fully formal) – ideally with value added at each step.<sup>15</sup>

The economics of governance responds to these injunctions as follows: economizing on transaction costs is named as the main case (keep it simple); the logic of economic organization is that of discriminating alignment (as described in Section 4, below), the cognitive and self-interestedness attributes of human actors are described in plausible ways, thereby to disallow fanciful constructions, and the derivation and testing of refutable implications are continuously featured.

#### 4. Vertical Integration and the Theory of the Firm as Governance Structure

The economics of governance has its origins in the vertical integration decision, after which a series of related contractual phenomena are interpreted as variations on a theme. Vertical integration – or, in more symmetrical terms, the make-or-buy decision – is the obvious phenomenon with which to begin for several reasons. First, vertical integration is the issue on which Coase focused in his classic 1937 paper.<sup>16</sup> Second, Arrow's remark that that "the existence of vertical integration may suggest that the costs of operating competitive markets are not zero, as is usually assumed by our theoretical analysis" (1969, p. 48) concentrates transaction costs attention on vertical integration. Third, public

policy toward vertical integration in the 1960s was deeply confused. Finally, as compared with final product market transactions, vertical integration has the advantage of being much simpler, in that a variety of complications that arise in transactions between firms and consumers – such as disparities of information, differential access to technical and legal expertise, differential capacity to bear risk, and the like – are of second order importance in transactions between firms. The action in intermediate product market transactions thus resides in the attributes of transactions in relation to the properties of alternative modes of governance.

Coase, of course, is correct in his insistence that the boundary of the firm should be derived rather than taken as given. But where does the focal action reside? One possibility is that the differences in outside and internal procurement are to be discovered in the nature of the contract differences between a sales contract (outside procurement) and an employment contract (internal procurement) (Coase, 1937; Simon, 1951; Masten, 1988). Albeit instructive, the economics of governance proceeds differently. Taking as given that the workers within each technologically nonseparable stage are governed by an employment relation, the question is how to mediate the transfer between stages, especially as this has a bearing on adaptation.

This is not to suggest that these two contractual frames are disjunct. To the contrary, they share a good deal of common ground and it is easy to see how some viewed the employment relation as the cutting edge. The main advantages that I associate with the latter approach are these: (1) focusing on transactions

in intermediate product markets – make or buy – rather than on sales and employment relations calls attention to the need to identify the critical dimensions for describing transactions; (2) as described above, intermediate product market transactions pose fewer trading disparities between the parties than do employment relations and final product market transactions; (3) intermediate product market transactions are less emotive than is the employment relation; and (4) a wider range of governance alternatives, to include ownership of both stages, comes under scrutiny. It is also noteworthy that whereas interpreting many other contractual phenomena as variations on a vertical integration theme has taken shape easily, the employment relation has not witnessed similar progress.

The upshot is that the choice of vertical integration as the paradigm problem has had strategic advantages. I begin with the discriminating alignment hypothesis, which leads naturally into the dimensionalization of transactions and of governance structures. Many of the key intuitions for efficient governance are thereafter captured in the “simple contractual schema,” after which I briefly discuss several reduced form models.

#### 4.1 the discriminating alignment hypothesis

The analytical challenge is to deploy governance structures so as to organize transactions in an efficient way. The action devolves on the attributes of transactions and governance structures. The core argument is this:

- (1) If some transactions are simple and others are complex, then the attributes of transactions that are responsible for these differences must be named and their ramifications set out. The economics of governance responds by naming asset specificity (which can take a variety of forms), uncertainty, and frequency as three of the critical dimensions for describing transactions.
- (2) If the comparative efficacy of different modes of governance (market, hybrid, hierarchy, public bureau, etc.) differ, then the critical attributes that describe alternative modes of governance need to be named and the internally consistent syndromes of attributes that define viable modes need to be worked out.
- (3) A predictive theory of economic organization resides in the discriminating alignment hypothesis: transactions, which differ in their attributes, are aligned with governance structures, which differ in their costs and competencies, so as to effect a (mainly) transaction cost economizing result, which is the main case.<sup>17</sup>

#### 4.2 dimensionalizing transactions and governance

The critical attributes of transactions and the internally consistent syndromes of attributes that define viable modes of governance are discussed elsewhere (Williamson, 1979, 1985, 1991). I merely offer some summary comments on both.

- (c) transactions

First, there is general agreement that asset specificity, uncertainty, and frequency are relevant dimensions for describing transactions. Much of explanatory power of the theory turns on asset specificity, which gives rise to bilateral dependency (or the absence thereof). Bilateral dependency by itself, however, would not pose a problem but for disturbances that push the parties to an incomplete contract off of the contract curve. Indeed, the problem of contracting under fully stationary conditions is uninteresting: “Only when the need to make unprogrammed adaptations [to disturbances] is introduced does the market versus internal organization issue become engaging” (Williamson, 1971). Frequency is relevant in two respects: reputation effects and setup costs. Higher frequency transactions support experience rating, to which reputation effects can thereafter be applied. Greater frequency also permits the cost of specialized management and dispute resolution machinery to be amortized more easily. The net effects will vary with the particulars, especially with respect to the degree of confidence that is ascribed to reputation effects (the limited efficacy of which often goes unremarked).

(b) alternative modes of governance

Governance focuses on ongoing contractual relations and describes alternative modes of governance in discrete structural terms.

The basic setup distinguishes among three modes of governance: classical markets, hybrid modes of contracting, and hierarchy. The three attributes for describing governance structures are incentive intensity,

administrative controls, and contract law regime. Simplification is accomplished by allowing each attribute to take on one of three values: much (++), some (+), or little (0). There being three attributes that can take on three values, 27 different combinations can be described. What are the defining attributes of markets, hybrids, and hierarchies?

The intuition that hybrids are somehow located between markets and hierarchies is widely shared. On the provisional assumption that this is correct, it will be instructive to begin with an examination of how markets and hierarchies differ in incentive, control, and contract law respects.<sup>18</sup>

With reference to incentives, the argument is that markets work out of high-powered incentives, in that each party is autonomous and appropriates its net receipts. Interfirm controls in market exchange are mainly limited to inspections of quantity and quality at the point of delivery. Otherwise each firm independently manages its accounting and auditing; and adaptations to disturbances are likewise of an autonomous kind. Disputes, moreover, are settled in a legalistic way, with emphasis on money damages, since neither side has a strong economic interest in preserving the continuity of the relationship.

Incentives in hierarchy, by contrast, are comparatively low-powered – which is both an unavoidable consequence of unified ownership and, more to the point, an intentional feature of hierarchy. If coordinated adaptation in the *raison d'être* of internal organization, and if the owner of an enterprise possesses the authority to decide when and how to intervene if disturbances push an internal procurement relation out of alignment, then any “promise” of autonomy to a

subordinate division is a delusion. (The issues here relate to the impossibility of selective intervention, as discussed above.) Suffice it to observe here that the move from market to unified ownership (hierarchy), thereby to promote coordinated adaptation, is always attended by a reduction in incentive intensity.

Because weaker incentive intensity has manifold consequences, of which malingering is one, a variety of procedural controls (rules and regulations) are put in place. Such controls have both immediate cost and future bureaucratic cost consequences. In consideration of these added costs, transactions for which coordinated adaptations are unneeded or promise only small gains will remain in the market rather than be moved into hierarchy.

A further feature of hierarchy that operates in the service of (uncontested) coordinated adaptations is that the firm becomes its own court of ultimate appeal, as supported by forbearance law. Taken together, the comparison of markets and hierarchies in governance respects comes down to the following:

<u>Instrument</u>	<u>Market</u>	<u>Hierarchy</u>
Incentive:	high powered	low powered
Controls:	limited	extensive
Contract Law:	legalistic	forbearance

So much for the polar cases. What about hybrids? The intuition here is that hybrids are used to manage transactions where the degree of asset

specificity and the nature of the disturbances to which they are subject are of “intermediate degree.” The putative attraction of hybrids is that (1) as compared with markets they are better able to infuse order, thereby to mitigate conflict and realize mutual gains for transactions where continuity matters yet (2) they avoid the heavy bureaucratic cost burdens of hierarchy.

The viability of hybrid contracting rests heavily on the concept of “credible commitment.” Upon recognizing that transactions that are supported by investments in specific assets pose a hazard of bilateral dependency, parties to such transactions have incentives to craft the aforementioned interfirm contractual safeguards (penalties against premature breach; information disclosure and verification; dispute settlement mechanisms) that promote contractual continuity. Llewellyn’s concept of contract as framework, as against legal rules, applies. The contract law of hybrids is thus located between that of market (legalistic) and hierarchy (forbearance).

Also, by comparison with markets, the hybrid modes sacrifices incentive intensity, but not to the degree of hierarchy. And hybrid modes employ more administrative controls than markets, but not in the degree of hierarchy. The syndromes of attributes that describe the internally consistent attributes that are associated with markets, hybrids, and hierarchies are thus as shown in Table 1.

	----- Governance Structure -----		
<u>Attributes</u>	Market	Hybrid	Hierarchy
Incentive Intensity	High-powered	Semi-high powered	Low-powered

	(++)	(+)	( 0 )
Administrative Controls	little	some	much
	( 0 )	(+)	(++)
Contract Law Regime	legalistic	contract as framework	forbearance
	(++)	(+)	( 0 )
<b>Table 1. Attributes of Three Viable Modes of Governance</b>			

(c) efficient alignment

As discussed earlier, the economics of governance takes adaptation to be the central purpose of economic organization and makes provision for both autonomous and coordinated adaptations. The syndrome of attributes that describe the market is especially well-suited to manage transactions where the adaptive needs are principally those of autonomous adaptation, while the syndrome that describes hierarchy is well-suited to support cooperative adaptation between successive bilaterally dependent stages. The hybrid mode is a compromise syndrome while does well, but not surpassingly well, in both autonomous and coordinated adaptation respects. Which mode to use when depends, therefore, on the mix of adaptive needs posed by the transactions under examination (Williamson, 1991).

#### 4.2 the simple contractual schema

Assume that a firm can make or buy a component and assume further that the component can be supplied by either a general purpose technology or a special purpose technology, where  $k$  is a measure of asset specificity. The transactions in Figure 1 that use the general purpose technology are ones for which  $k = 0$ . In this case, no specific assets are involved and the parties are essentially faceless. Those transactions that use the special purpose technology are ones for which  $k > 0$ . As earlier discussed, bilaterally dependent parties have incentives to promote continuity and safeguard their specific investments. Let  $s$  denote the magnitude of any such safeguards, which include penalties, information disclosure and verification procedures, specialized dispute resolution (such as arbitration) and, in the limit, integration of the two stages under unified ownership. An  $s = 0$  condition is one for which no safeguards are provided; a decision to provide safeguards is reflected by an  $s > 0$  result.

Node A in Figure 1 corresponds to the ideal transaction in law and economics: there being an absence of dependency, governance is accomplished through competitive market prices and, in the event of disputes, by court awarded damages.<sup>19</sup> Node B poses unrelieved contractual hazards, in that specialized investments are exposed ( $k > 0$ ) for which no safeguards ( $s = 0$ ) have been provided. Such hazards will be recognized by farsighted players, who will price out the implied risks of contractual breakdown.

Added contractual supports ( $s > 0$ ) are provided at nodes C and D. At node C, these contractual supports take the form of interfirm contractual

safeguards. Should, however, costly contractual breakdowns continue in the face of best bilateral efforts to craft cost-effective safeguards at node C, the transaction may be taken out of the market and organized under unified ownership (vertical integration) instead. Because added bureaucratic costs accrue upon taking a transaction out of the market and organizing it internally, internal organization is usefully thought of as the organization form of last resort: try markets, try hybrids, and have recourse to the firm only when all else fails. Node D, the unified firm, thus comes in only as higher degrees of asset specificity and added uncertainty award priority to cooperative adaptation.

Note that the price that a supplier will bid to supply under node C (hybrid) conditions will be less than the price that will be bid at node B. That is because added security features ( $s > 0$ ) serve to reduce the risk at node C, as compared with node B, so the contractual hazard premium will be reduced. One implication is that suppliers do not need to petition buyers to provide safeguards. Because buyers will receive product on better terms (lower price) when added security is provided, buyers have the incentive to offer cost-effective credible commitments. Also note that whereas such commitments are sometimes thought of as a user-friendly way to contract, the analytical action resides in the hard-headed use of credibility to support those transactions where asset specificity and contractual hazards are at issue. Such supports are without purpose for transactions where the general purpose production technology is employed.

#### 4.3 reduced form models

Recall that (1) all complex contracts are incomplete (by reason of bounded rationality) and are subject to strategic breakdown (by reason of opportunism), where the probability of breakdown increases with asset specificity (bilateral dependency) and outlier disturbances,<sup>20</sup> (2) adaptation is the central problem of economic organization, of which two types are distinguished: autonomous adaptation to changes in relative prices (where the market enjoys the advantage) and cooperative adaptation (where the advantage accrues to hierarchy), (3) internal procurement is disadvantaged by the added burdens of bureaucracy, and (4) in deference to the precept “keep it simple,” the basic analysis turns entirely on the attributes of transactions in relation to the syndromes of attributes that define generic modes of governance (market, hybrid, and hierarchy).

Reduced form models into which these features have been introduced include:

(1) the credible commitment model comparing markets and hybrids, in which hostages are used to support otherwise hazardous exchange (Williamson, 1983), (2) a neoclassical variant comparing markets and hierarchies in which both production costs and transaction costs appear (Riordan and Williamson, 1985),<sup>21</sup> and (3) a stochastic version of the argument in which adaptation trade-offs are examined in a three-way comparison of markets, hybrids, and hierarchies (Williamson, 1991). Individually and collectively, these models capture the key intuitions of the simple contractual schema yet sometimes go beyond, in that they uncover nuances and identify critical parameters that are not picked up by the schema.

## 5. Beyond Vertical Integration

The vertical integration problem is a window upon the governance of contractual relations more generally, to include variations on a theme, public policy ramifications, and empirical testing.

### 5.1 Variations on a Theme

New paradigms have the property that a wide range of phenomena are informed by the same underlying logic (Kuhn, 1970, pp. 10, 23, 175), whereupon “superficially disconnected and diverse phenomena [are revealed] to be manifestations of a fundamental and relatively simple structure” (Friedman, 1953, p. 33). As the growing literature on complex contracting reveals, contractual regularities that are uncovered by an examination of vertical integration recur, with variation, across a wide array of contractual phenomena. Indeed, any issue that arises as or can be reformulated as a contracting problem can be examined to advantage in transaction cost economizing terms.

Recurring regularities of a generic kind include the attributes of human actors, intertemporal process transformations, and implicit contract law differences among alternative modes of governance. Thus all complex contracts are unavoidably incomplete by reason of bounded rationality; the preservation of ongoing contractual relations is important in the degree to which bilateral dependency sets in, in that such contracts are subject to defection from the spirit of cooperation, by reason of opportunism, when adaptations to outlier disturbances are needed to restore a position on the shifting contract curve; and

contract law differences, of which forbearance law is an example, contribute to the discrete structural differences among alternative modes of governance.

Discriminating alignment, moreover, tracks the logic of the simple contractual schema, albeit with variation. With respect to labor, for example, node D in Figure 1 drops out of the schema (because the firm is unable to own labor). The comparison therefore comes down to nodes A, B and C. Node A corresponds to the case where labor is easily redeployed to other uses or users without loss of productive value ( $k = 0$ ). Thus, although such labor may be highly skilled (as with many professionals), the lack of firm specificity means that, transition costs aside, neither worker nor firm has an interest in crafting penalties for unwanted quits/terminations or otherwise creating costly internal labor markets (ports of entry, promotion ladders), costly information disclosure and verification procedures, and costly firm-specific dispute settlement machinery. The mutual benefits do not warrant the costs.

Conditions change when  $k > 0$ , since workers who acquire firm-specific skills will lose value if prematurely terminated (and firms will incur added training costs if such employees quit). Here, as elsewhere, unrelieved hazards (as at node B) will result in demands by workers for a hazard premium, and recurrent contractual impasses, by reason of conflict, will result in inefficiency. Because continuity has value to both firm and worker, governance features that deter termination (severance pay) and quits (nonvested benefits) and that address and settle disputes in an orderly way (grievance systems) to which the parties ascribe confidence have a lot to recommend them. These can, but need not, take the

form of “unions.” Whatever the name, the object is to craft credible commitments (at node C) in which the parties have mutual confidence and that enhance efficiency (Baron and Kreps, 1999, pp. 130-138; Williamson, 1975, pp. 27-80, 1985, pp. 250-262).<sup>22</sup>

Or consider the distinctions between debt and equity. Viewed through the lens of contract/governance, the board of directors is interpreted as a security feature that arises in support of the contract for equity finance. More generally, debt and equity are not merely alternative modes of finance, which is the law and economics construction (Easterbrook and Fischel, 1986; Posner, 1986), but are also alternative modes of governance. In project finance terms, debt (which is akin to the market form) is more well-suited to finance generic investments whereas equity (which is the more administrative form) is used to finance more specialized investments (Williamson, 1988).

These and other variations on a theme<sup>23</sup> notwithstanding, it bears emphasis that nuanced understandings of puzzling contractual practices (such as exchange agreements, whereby petroleum firms exchange rather than sell product to one another, is an illustration (Williamson, 1985, Chap. 8)) often turn on deep knowledge of the phenomena. Credible contracting, for example, takes a variety of forms, the particulars of which vary with the specifics of the good or service in question and the industry of which it is a part (sometimes to include path dependence features). Thus although the overarching framework that arises out of the study of vertical integration has broad application, the

interpretation of particulars often requires transaction-specific knowledge of a more detailed and nuanced kind. The action, to repeat, is in the microanalytics.

## 5.2 public policy

All-purpose reliance on neoclassical economic reasoning contributed significantly to the developing crises in antitrust enforcement and regulation in the 1960s. Indeed, interest in the economics of governance was stimulated in part by overuse of monopoly reasoning in the antitrust arena, where non-standard contractual practices and organizational forms were presumed to have monopoly purpose and effect (Coase, 1972, p. 67):

One important result of this preoccupation with the monopoly problem is that 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 understandable practices tends to be very large, and the reliance on a monopoly explanation, frequent.

Lacking an appreciation for the efficiency benefits of governance, nonstandard forms of contracting were interpreted not “hospitably, in the common law tradition, but inhospitably in the tradition of antitrust.”<sup>24</sup> One striking illustration of this state of affairs is the muddled economic reasoning in the jurisdictional statement and brief for the government in the Schwinn case (which required an interpretation of economic purposes served by vertical market

restrictions imposed on Schwinn franchisees) in arguments presented to the U.S. Supreme Court in 1967.<sup>25</sup> Another illustration is the confusion that arose over merger, or other extensions to the boundary of the firm, beyond those that could be justified by economies of scale or scope. Because these were believed to have monopoly purpose and effect, mergers with even tiny market shares were subject to antitrust challenge. (As Justice Stewart put it, in a dissenting opinion in 1966, the “sole consistency that I can find is that in the {merger} litigation under Section 7, the Government always wins.”)<sup>27</sup> The propensity to recommend regulation as a “solution” to market failures, without symmetrical regard for the failures – disincentives, costs, distortions, misconceptions – to which regulation was subject (Coase, 1964), likewise led to excesses in regulation.

The lens of contract/governance approach to public policy helped to restore perspectives in antitrust and regulation and more generally. As Dixit remarks, the era of black box applied welfare economics had left (1996, p. 9):

... some very important gaps in our understanding and [gave] us some very misleading ideas about the possibilities of beneficial policy intervention. Economists studying business and industrial organization have long recognized the inadequacy of the neoclassical view of the firm and have developed richer paradigms and models based on the concepts of various kinds of transaction costs. Policy analysis also stands to benefit from ... opening the black box and examining the actual workings of the mechanism inside.

Rather than continue the uncritical practice of comparing actual alternatives with hypothetical ideals, attention needed to be focused instead on comparing feasible alternatives, all of which are flawed. Lapses into reasoning of an omniscient, omnipotent, benevolent kind is avoided by (1) recognizing that it is impossible to do better than one's best, (2) insisting that all the finalists in a governance structure competition meet the test of feasibility, (3) symmetrically exposing the strengths and weaknesses of all proposed feasible forms, and (4) describing and costing out the mechanisms of implementation. As against hypothetical ideals of a zero transaction cost kind, the "remediableness criterion" holds that an extant mode of governance for which no superior feasible mode can be described and implemented with expected net gains is presumed to be efficient. Feasibility, implementation costs, and the rebuttable presumption of efficiency (which can bring politics into play) are thus the key features in a constructive effort to reshape the way that public policy analysis gets done.

### 5.3 empirical testing

By comparison with most economic theories of organization, the economics of governance is an empirical success story.<sup>27</sup> As of the year 2000 there were over 600 empirical articles on transaction cost economics with exponential growth therein (Boerner and Macher, 2001).<sup>28</sup> There is no occasion to rest content, however. All theories of economic organization, the economics of

governance included, will benefit from more and better empirical tests: better data, additional phenomena, better statistical procedures.

It is nonetheless noteworthy that much of the empirical research on the economics of governance does not rely on published data that have been collected for an altogether different purpose (e.g., to satisfy census or regulatory requirements). Instead, most of the best empirical research on governance uses primary data that have been collected with the microanalytic needs of the discriminating alignment hypothesis foremost in mind.

Interestingly, Bengt Holmstrom and Jean Tirole conclude their survey of the theory of the firm literature with the observation that “the evidence/theory ratio ... is currently very low” (1989. p. 126) and recommend that this be rectified. But might this low ratio be explained more by the shortage of predictions than it is by nonresponsiveness among empirically minded social scientists, who are continuously attuned to testable, refutable implications? In that event, theoretical contributors to the theory of the firm literature are encouraged to award priority to the derivation of testable, refutable implications (Whinston, 2003), thereby to improve the evidence/theory ratio.

## 6. Tensions and Challenges

The economics of governance is a work-in-progress. Among the issues that warrant further work are scaling up, the challenges posed by doing contracting and organization in high velocity environments, the uses of game theory (and full formalism more generally), and extensions upon the basic setup.

## 6.1 scaling up

Most contractual theories of the firm work out of a microcosm, often the individual make-or-buy decision. Few such theories raise the issue of scaling up, and fewer still address it – either at all or in a plausible way. One possible explanation is that because all of the key insights are revealed by the simple model, scaling up is unneeded. As recent applications of simple models attest, however, that is unduly sanguine: scaling up is commonly attended by organizational complications. Given our current primitive level of understanding, such complications need to be uncovered and confronted, lest simple models lead to avoidable error.

The economics of governance responds to the challenge of scaling up by applying the logic of efficient alignment to a series of technologically separable stages of activity. Broadly in the spirit of the “architecture of complexity” (Simon, 1996, Chap. 8), the argument is that some technologically separable stages should remain freestanding (in the market) but others should be brought within the boundary of the firm, depending on the need for coordinated adaptations at the interface. Market procurement is appropriate for technologically separable stages that produce generic goods and services, since these have little need for coordinated adaptations. Successive stages, however, that are bilaterally dependent are taken within the boundary of the firm and managed in a two-part way: small and frequent local disturbances within a stage are managed by the application of standard operating procedures of “routines” (Cyert and March,

1963), whereas more consequential disturbances that span subsystems warrant “management by exception,” hence are coordinated by hierarchy.

To be sure, cumulative bureaucratic cost burdens accrue as successive transactions are folded in, so interaction effects among transactions that are technologically separable occur in this way. As a first approximation, however, the boundary of the firm (or, perhaps, of an operating division within a multi-divisional structure (Chandler, 1966)) is defined by that subset of technologically separable transactions for which coordinated adaptations yield net gains. Albeit crude, such an approach constitutes a beginning and compares well with other contractual theories of the firm in scaling up respects.<sup>29</sup>

## 6.2 real time

The economics of governance assumes that parties to transactions have the time to make the requisite investments in transaction specific assets in an unhurried, deliberative way. There are some transactions, however, for which real-time responsiveness is of the essence. What may be referred to as T-form structures, where T denotes temporary or transitional modes for which timeliness matters, can play a huge role in the success or failure of firms that are operating in newly developing markets where technology and rivalry are undergoing rapid change (Williamson, 1995, pp. 40-42).

Some joint ventures, alliances, and research consortia are of this kind, in that they are responsive to exigencies where each party is unable, by itself, to assemble and deploy the requisite resources in a timely way. By reason of

urgency (as well as out of respect for the added burdens of bureaucracy that would attend an inclusive merger of the relevant stages), the requisite resources are assembled instead by pooling.

Describing joint ventures, alliances, and research consortia that owe their origins to exigencies as disequilibrium T-form structures (rather than as equilibrium hybrids) may help to explain the puzzle that both successful and unsuccessful T-forms are commonly terminated when contracts expire. Successful T-forms will be terminated because the combined effort has permitted each party to remain viable and learn enough and/or buy time to go it alone. Unsuccessful T-forms will be terminated because the opportunity to participate will have passed them by.<sup>30</sup> A deeper understanding of T-forms may well require the development of new conceptual apparatus (Aoki, 2004), some of it a game theoretic kind. As matters stand presently, understanding T-form structures is as yet an unmet challenge.

### 6.3 game theory

David Kreps views the relation between game theory and the economics of governance as reciprocal: not only have “important ideas in Markets and Hierarchies have been translated into the more mathematical language of non-cooperative game theory,” of which more such is in prospect (1999, p. 122), but the relation is reciprocal: game theory “will [also] give something back” (1999, p. 122).<sup>31</sup> Five years have expired since and applications of game theory to the study of contract have grown apace. To be sure, many of these are of an agency

theory (ex ante incentive alignment) kind, but governance is also receiving more formal modeling attention.

Dixit proposes two criteria for assessing the merits of game theoretic models of governance: “First, a model should do more than explain just the simplest and most obvious facts ..., it should offer a sharper or deeper understanding of them in terms of more basic economic principles or by connecting them to other seemingly unrelated facts. Second, a theoretical model should not merely reproduce as results the factual results of case studies that the model was constructed to explain in the first place; it should yield some new results or hypotheses that can be compared with other facts” (2004, p. 22).

Although I am persuaded that tool users who abide by these criteria will help us to realize the elusive goal of fully formal models of governance, I would caution that parts of the economics of governance may continue to defy full formalization.<sup>32</sup>

#### 6.4 unfinished project

The limits of full formalization aside, the economics of governance stands to benefit from further work of a semi-formal kind. Attributes of transactions other than those of asset specificity, uncertainty, and frequency may have explanatory value for describing some transactions. Also, governance structures of nonprofit, political, regulatory, and governmental kinds are obvious candidates for analysis for which additional governance dimensions may prove vital. And although

efficiency has broad sweep and economic importance, purposes other than efficiency may sometimes need to be factored in.

I would nevertheless recommend that such efforts proceed with caution. For one thing, efficiency reasoning can often subsume values such as loyalty and trust, especially as loyalty and trust are applied to transactions in a conditional (which is to say calculative) way. Second, noncalculative applications of user-friendly terms such as trust, loyalty, altruism, and the like often result in truncated reasoning and lead to avoidable confusion (Williamson, 1996, Chap. 10).

## 7. Conclusions

The economics of governance joins to two fundamental concepts – governance and transaction costs – with the purpose of pouring operational content into both. By contrast with orthodoxy, where continuity of the exchange relation is unimportant, the economics of governance treats simple market exchange as a special case and focuses predominantly on those transactions for which continuity of the exchange relation is valued, even vital. There being many economic transactions that qualify, the economics of governance has broad reach.

To be sure, the economics of governance also entails set-up costs. Students of governance need to examine economic phenomena through a different lens (the lens of contract) than orthodoxy (the lens of choice), be prepared to adopt a more interdisciplinary orientation (in which law, economics, and organization theory are selectively combined), and uncover and examine the

mechanisms of governance in relation to the attributes of transactions in a more microanalytic way than had been customary in economics (Arrow, 1987). Is the game worth the candle?

I venture a two-part answer. First, not everyone will want to or should make the requisite investments. The economics of governance, after all, is only one of several instructive lenses for studying the economics of organization. But second, I submit that our understanding of economic organization, and public policy pertinent thereto, has been impoverished by failures to pay heed to the lessons of governance. If and as others agree, the economics of governance is an unfinished project whose time has come.

### **Footnotes**

- \* The author is Professor of the Graduate School and Edgar F. Kaiser Emeritus Professor of Business, Economics, and Law at the University of California, Berkeley.
- 1. Whereas contract economics concentrates all of the analytical action in the (ex ante) incentive alignment stage of contracting, governance is predominantly concerned with the management of ongoing contractual relations during the (ex post) contract implementation interval.
- 2. Such pluralism is warranted partly because the social sciences “deal with phenomena of the greatest complexity” (Simon, 1957a, p. 89). “The social sciences are hyper-complex. They are inherently far more difficult than physics and chemistry” (Wilson, 1998, p. 183).

3. Governance costs focus principally on the costs of adaptation/maladaptation during the contract implementation interval (Williamson, 1985, p. 21). Such costs are a subset of transaction costs, which, broadly construed, are the “costs of running the economic system” (Arrow, 1969, p. 48).
4. Students of the history of thought will remind us that catallactics – meaning “the science of exchanges” – has much earlier origins. Indeed, a book by E. B. de Condillac on this subject was published in 1776, which is when The Wealth of Nations first appeared (see Murray Rothbard (1987, pp. 377-378) for an historical sketch). Recurrent interest in the science of contract notwithstanding, it has operated in the shadows of the science of choice. Why the disparity? Here as elsewhere, good ideas need to be operationalized.
5. The quote originates with John Schlegel (1979, p. 459), who used it to describe the fate of American Legal Realism. A similar fate befell American Institutional Economics, of which Commons was a leader.
6. Commons turned instead to W. N. Hohfeld’s system of “fundamental legal concepts” to implement his ideas. (For his reliance on Hohfeld, see especially the extended footnote in Commons (1968 [1924], p. 91).) The resulting effort to interpret transactions and ongoing concerns with the use of juridical reasoning resulted in an elaborate taxonomy (Commons, 1968 [1924], pp. 90-142), but a predictive theory of contract and organization and a follow-on empirical research agenda did not materialize.

7. On this, see Coase (1988, 1992), Douglass North (2000), Lars Werin (2000), Oliver Williamson (2000a), and Claude Menard (2000).
8. Armen Alchian and Harold Demsetz (1972) also made note of the tautological reputation of transaction cost economics, whereupon they focused their attention on technological nonseparabilities, as these relate to the “shirking problem” in team production, as the way by which to move the concept of transaction cost in a more operational direction. Their arguments that “the firm ... has no power of fiat, no authority, no disciplinary action any different from ordinary market contracting” and that “long-term contracts between employer and employee are not the essence of the organization that we call a firm” (1972, p. 777) are more in the spirit of agency theory than of the economics of governance.
9. Whether Commons or Coase would agree on the merits of such a linkage is unclear, especially as Coase was dismissive of the contributions of American Institutionalists, Commons presumably included (Coase, 1984, p. 230).
10. Economics and the contiguous social sciences, especially economics and sociology, mainly went their own ways during the 1940s and 1950s. Thus, Paul Samuelson (1947) distinguished between economics and sociology in terms of their rationality orientations, with rationality being the domain of economics and nonrationality being relegated to sociology. James Duesenberry subsequently quipped (1960) that economics was preoccupied with how individuals made choices, whereas sociology maintained that

individuals did not have any choices to make. Slowly but surely, the yawning disjunction between economics and the contiguous social sciences would change.

11. Interestingly, Jean Jacques Laffont and David Martimort (2002) also credit Barnard as “the first to define a general theory of incentives in management” (2002, p. 11), where they interpret Barnard’s views as broadly in the spirit of their own agency theory work. Laffont and Martimort also write that “Barnard recognized that incentive contracts do not rule all of the activities within an organization.” In particular, “the incompleteness of contracts and the bounded rationality of members of the organization require that some leaders be given authority” (2002, p. 13), presumably to exercise ex post governance.
12. Another class of intertemporal transformations that have organization design ramifications are the “informal organization” consequences that Barnard associates with formal organization (1938, p. 20), which can be interpreted in part as “corporate culture” (Kreps, 1990). The basic informal organization insight is this: taking transactions out of markets and organizing them internally subjects them to different intertemporal governance influences, which effects should be uncovered and factored into the calculus in deciding on the relative merits of markets and the hierarchies.
13. The accommodation to which I refer is related to Barnard’s observation that “Perhaps often and certainly occasionally, men cannot be promoted or

selected, or even must be relieved, because they cannot function, because they 'do not fit,' where there is no question of formal competence" (1938, p. 224).

14. Some nevertheless favor "insight" over prediction as the cutting edge:  
"Faced with a choice between a theory which predicts well but gives us little insight into how the system works and one which gives us this insight but predicts badly, I would choose the latter, and I am inclined to think that most economists would do the same" (Coase, 1994, p. 17). Indeed Coase contends that "economists do not, could not, and, if they could, should not choose their theories on the basis of the accuracy of their predictions" (1988, p. 24).

Although I am persuaded that understanding is what we are up to, I am loathe to judge theories in terms of "insight." The basic problem is that insights vary with predilections and tastes and are easily politicized:

"Science, viewed as a competition among theories, has an unmatched advantage over all other forms of intellectual competition. In the long run (no more than centuries), the winner succeeds not by superior rhetoric, not by the ability to convince or dazzle a lay audience, not by political influence, but by the support of data, facts as they are gradually and cumulatively revealed. As long as its factual veridicality is unchallenged, one can remain calm about the future of a theory" (Simon, 1991, pp. 364-365).

15. For a recent review and discussion of fully formal models of economic organization, see Robert Gibbons (2005).

16. A still earlier precedent is suggested by Adam Smith's celebrated description of the division of labor in pin manufacture (1776), which can be interpreted as a succession of vertically integrated stages of manufacture.
17. This not to say that technology and monopoly are irrelevant, but neither qualifies as the main case.
18. As Harold Demsetz observes, 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" (Demsetz, 1983, p. 377). The chief mission of the economics of governance is to understand the powers and limits of alternative modes of governance, the inner workings of real firms included, and to align these with the adaptive needs of different transactions.
19. Transition economies with primitive courts rely more on private ordering (McMillan and Woodruff, 1999).
20. The basic proposition here is that when the "lawful" gains to be had by insistence upon literal enforcement exceed the discounted value of continuing the exchange relation, defection from the cooperative spirit of the contract can be anticipated (Williamson, 1991, p. 273; Klein, 1996).
21. This is responsive to the observation that new theories of economic organization are believed to take on greater "analytical usefulness when these are founded on more directly neoclassical lines" (Arrow, 1985, p. 303).

22. The emphasis on collective organization as a governance response is to be distinguished from the earlier work of Gary Becker, where human asset specificity is responsible for upward-sloping age-earnings profiles (Becker, 1962). Becker's treatment is more in the science of choice tradition, whereas mine views asset specificity through the lens of contract. These two are not mutually exclusive. They do, however, point to different empirical research agenda.
23. Other phenomena that can be interpreted as variations on a theme include vertical market restrictions, to include franchising and forward integration into distribution, the uses of regulation and deregulation, the order of privatizing industries in economies that are undergoing transition, reputation effects (and the limits therein), the provision of consumer and worker health and safety protections (especially to hazards with long latency effects), multinational economic organization, corporate strategy, and the list goes on. For collections of articles that work out of the logic of governance, see Williamson and Masten (1995) and Claude Menard (2005), especially those articles that Menard clusters under headings VII through XIII. Also note that, upon replacing  $k$  (which denotes asset specificity) by  $h$  (which denotes contractual hazards more generally), describing the attributes of additional modes of governance (nonprofits, government bureaus, cooperatives, etc.), and invoking the discriminating alignment hypothesis, the spirit of the simple contracting schema can be extended to include a wide range of non-commercial contracts and organizations.

24. The quoted language is that of the then head of the Antitrust Division, Donald Turner. See Stanley Robinson, 1968, N.Y. State Bar Association, Antitrust Symposium, p. 29.
25. For a discussion, see Williamson (1985, pp. 183-184).
26. United States v. Von's Grocery Co., 384 U.S. 270, 301 (1966) (Stewart J. dissenting).
27. As Scott Masten observes, "surveys of the empirical transaction cost literature attest...[that] the theory and evidence have displayed remarkable congruity" (1995, pp. xi-xii). In Paul Joskow's judgment, moreover, empirical work of a transaction cost kind was "in much better shape than much of the work in industrial organization generally" (1991, p. 81).
28. To be sure, most of the empirical studies take the attributes of the transaction as given when in fact the choice of transaction attributes and governance structures is done simultaneously (Riordan and Williamson, 1985). Checks on simultaneous equation bias nevertheless suggest that any such effects are small. Repeated findings of corroboration for the predicted association between transaction attributes and organization form across a wide variety of phenomena are not spurious. Those who have done the modest, slow, molecular, definitive empirical work in this area deserve enormous credit.
29. The Jensen and Meckling informal agency theory model was designed to ascertain the performance consequences of selling equity in an owner-managed (entrepreneurial) firm. The model sheds insights into the

immediate consequences of dilution as well as cost-effective ways of mitigating these effects. However, the real interest of Jensen and Meckling (and of the corporate governance literature that was influenced by this path breaking paper) was not in entrepreneurial firms but in the modern corporation. Jensen and Meckling thus admit to the following unmet need (1976, p.356):

One of the most serious limitations of this analysis is that as it stands we have not worked out in this paper its application to the very large modern corporation whose managers own little or no equity. We believe our approach can be applied to this case but ... [these issues] remain to be worked out in detail and will be included in a future paper.

Their conjecture that such scaling up can be accomplished may turn out to be correct, but this part of the project has yet to be “worked out in detail” – by them or others.

30. Note, however, that not all joint ventures, alliances, and research consortia have their origins in real-time responsiveness. Some are organizational solutions to enduring problems. And some are mistaken ventures.
31. Speaking “as a tool fashioner interested in developing tools that deal better with the world-as-it-is,” Kreps avers that “game theory (the tool) has more to learn from transaction cost economics than it will have to give back” (1999, p. 122). The ratio of “give back” to “take” has been increasing.

32. I am not qualified to say whether a new mathematics will be required, but that possibility has been called to my attention. Thus Oskar Morgenstern, in his “Prolegomena to a Theory of Organization,” remarked on the need “to establish a language, reasonably accurate yet rich enough to provide the possibility of empirical description that must precede the formulation of any [fully formal] theory. The mathematical character of the future theory may be indiscernible at present. It may well happen, as stated in the Theory of Games and Economic Behavior ... that fundamental new mathematical discoveries comparable to the creation of calculus may have to be made in this field” (1951, p. 11). Also, a then-young mathematical economist colleague of mine ventured the opinion in the early 1980s that we would need a new mathematics to work out the fully formal logic of complex economic organization. The disciplined study of economic organization being very daunting, I can appreciate his concerns. But others are much better qualified than I to make the determination.