

**Transaction Cost Economics:
What Are the Questions?**

Oliver E. Williamson
University of California, Berkeley

April, 2012

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

Transaction Cost Economics: What Are the Questions?

Oliver E. Williamson

The origins of Transaction Cost Economics can be variously dated. Here and elsewhere I trace its origins to path breaking developments in law, economics, and organization theory in the 1930s. My participation in the project began with my paper “The Vertical Integration of Production: Market Failure Considerations“ (1971). Persuaded as I was that (1) price theoretic explanations for vertical integration that rested on efficient factor proportions and monopoly power abuses were too narrow and (2) that organization is important and is susceptible to analysis, my purpose was to ascertain if and how vertical integration relieves outsourcing problems. Specifically, I employed a comparative contractual approach to ascertain the differential transaction cost consequences of organizing different types of transactions in firms and markets.¹

Although I initially viewed this paper as a stand-alone project, I subsequently discovered that any issue that arises as or can be reformulated as a contracting problem can be examined to advantage in transaction cost economizing terms. So construed, the applications are endless.

I do not, however, mean to suggest that Transaction Cost Economics (TCE) is an all-purpose construction. Inasmuch as economics (and the social sciences more generally) are very complex as compared with the natural and physical sciences (Simon, 1957, p. 89; Wilson, 1998, p. 183), pluralism has a great deal to recommend it (Simon,

1992, p. 21). Complex social science phenomena are thus usefully examined through several focused perspectives.

The study of complex phenomena commonly evolves in a gradual way. As Allen Newell puts it, “Theories cumulate... They are refined and reformulated, corrected and expanded ... Theories are things to be nurtured and changed and built up“(1990, p. 14). As between confidently pronouncing “This is the law here!” when examining a complex social science issue, TCE asks the question instead “What is going on here?”²

This paper describes the “natural progression“in TCE as it moves across informal to pre-formal to semi-formal and fully formal stages of analysis. By contrast with textbook economic theory, TCE is an exercise in positive economics where marginal analysis gives way to discrete structural analysis and the lens of choice is supplanted by the lens of contract (with emphasis on ex post governance rather than ex ante incentive alignment). It also describes human actors in more veridical terms, is selectively interdisciplinary, and rests its case on refutable implications and empirical testing. By contrast with older style institutional economics, TCE demonstrates that not only are institutions important but they are also susceptible to analysis.

Section 1 is an anecdote. Section 2 provides background on TCE from the 1930s through the 1960s. Section 3 describes the behavioral, contractual, adaptative, disciplinary, and methodological foundations on which TCE rests. Section 4 implements the basic TCE project. Section 5 deals with refinements. The conclusions appear in Section 6.

1. An Anecdote

I was attending a National Bureau of Economics Research conference on “The Analysis of Public Output“ in 1969 and was speaking with Kenneth Arrow at one of the breaks. Arrow looked around the room and recounted, with obvious pleasure, the names of former students of his that were in attendance. He then added, to my surprise, that I was also “one of his students”– which was technically correct, although I took only one course from him (as an elective from the Stanford MBA program that I was then enrolled in) and I would soon thereafter leave Stanford to complete my PhD training at the Graduate School of Industrial Administration at Carnegie-Mellon (on more of which later). Arrow said that he remembered me because I “asked good questions“ and furthermore stated that this was what brought him to the attention of his teachers when he was a graduate student.

I had occasion to relate this event to one of my graduate students who, I think, took Arrow's remark to be faint praise. I nevertheless regarded it, then and now, as an honor – both because of my deep respect for Ken and because I think that asking good questions is undervalued as compared with writing good examinations. Indeed, I conjecture that a major reason why many PhD students have difficulty in making the transition from excellent qualifying exams to getting a dissertation underway is because they have been preoccupied with the mathematics of the models to the neglect of curiosity – by stepping back, now and then, to ask the question “What is going on here?”

To be sure, asking that question is no guarantee that a productive research project on this subject will shortly thereafter take shape. There are many dry holes. Still, asking the question “What is going on here?” has a lot to recommend it – especially for students of economics with applied interests.

2. Background

This section deals with transaction cost economics over the period 1930 to 1970. I begin with the extraordinary decade of the 1930s, turn next to the period 1940-1960, and conclude with new developments in the 1960s, to include my experiences as a student, teacher, researcher, and as Special Economic Assistant to the head of the Antitrust Division.

2.1 The 1930s

The 1930s was an exceptional decade for economics and the contiguous social sciences as well as science more generally. My emphasis is on those new ideas that had or would have an important influence on transaction cost economics.

2.1.1 Economics

Ronald Coase's famous 1937 paper on "The Nature of the Firm" uncovered an unexamined lapse in the textbook theory of firm and market organization. Rather than take the decision to make-or-buy a good or service as given, which was the prevailing practice, this should be derived. In effect, economists were advised to pose the question "What is going on here?" with respect to (1937, p. 389):

... 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.

Coase, moreover, pushed beyond by asking two follow-on questions: What is the missing concept? And where is the basic action concentrated? Although he did not

expressly state that transaction cost was the missing concept, that was, in effect, the neglected issue to which he directed attention. In choosing between the employment relation and the intermediate product market transaction as the paradigm problem, the employment relation was the transaction to which he directed attention.

But for this last, this is a remarkably prescient start. Taking the employment relation, rather than the intermediate product market transaction, to be focal, however, would prove to be an obstacle to the development of TCE. For one thing, “Obvious peculiarities of the employment relation ... distinguish it from other kinds of contracts“ (Simon, 1951, p. 293), on which account it is difficult to interpret other types of transactions as variations on an employment relation theme. Also, as Coase would subsequently observe, his emphasis on the employer-employee “misdirects attention“ from core issues (1988, p. 37).

2.1.2 Organization theory

Chester Barnard came to the study of organization not as a social scientist but as a deeply perceptive practitioner who was persuaded that organization was important and was susceptible to analysis. Finding little in the social science literature that recognized the importance of “the processes of coordination and decision“ or that “formal organization was the most important characteristic of social life“ (1938, p. ix), Barnard set about to correct this state of affairs.

What interested Barnard was not adaptation in the market but rather adaptation of a “conscious, deliberate, purposeful kind“ within the firm (1938, p. 4). Of special relevance to Barnard in this connection were (1) the development of a theory of authority, (2) an understanding of the employment relation, (3) an appreciation for informal

organization as a support both for formal organization and the integrity of the individual, and (4) reliance on an economizing orientation. Barnard concluded his book with the observation that we need a “science of organization” (1938, p. 290) – but that would take shape gradually and is still a work in progress..

2.1.3 Law

As against the standard practice of there being one all-purpose law of contract, Karl 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 concept 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 work" (1931, pp. 736-737). This last condition is important, in that recourse to the courts for purposes of ultimate appeal serves to delimit threat positions. Introducing the concept of “contract as framework“ in support of on-going relations contemplates reciprocal cooperative responses by the parties to deal with unanticipated disturbances, which differs consequentially from 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.)

2.2 1940-1960

2.2.1 Economics

Coase (1937) was essentially ignored over the interval 1940-1960 – partly because its relevance did not register and partly because so many other things that were perceived to be of greater importance were in progress. Keynesian economics is the obvious example.

Monopolistic competition was believed to be both more important and more tractable. The Socialist Controversy was a huge distraction. Mathematical economics was rapidly developing

and was the wave of the future. Not only were economics and organization theory viewed as oil to water (Samuelson, 1947; Duesenberry, 1960), but the concept of transaction cost was so broad that it could be used as an ex post rationalization for any outcome whatsoever.³ Awaiting a demonstration that the concept had teeth, better to ignore or dismiss it.

2.2.2 Organization theory

Organization theory did, however, progress during this interval. Simon was actively involved in four of them: his 1947 book Administrative Behavior relied extensively upon and called attention to Barnard's path breaking contributions; his 1952 paper in *Econometrica* on a "Formal Theory of the Employment Relation" was an effort at full formalism; his book (with James March) on Organizations (1958) was an ambitious effort to interpret and advance the organization theory literature; and Simon had a central role as a member of the faculty of the newly organized Graduate School of Industrial Organization at Carnegie-Mellon, where a group of young turks⁴ transformed training in the business schools (Angier and March, 2011) and promoted interdisciplinary social science research.

Contributions by sociologists to the study of organization include Philip Selznick's observation that organization, like the law, has a "life of its own" (1950, p. 10), in which case intertemporal regularities of organization need to be identified and the ramifications worked out. Also Robert Merton (1936) and Alvin Gouldner (1950) called attention to the "unanticipated consequences" that attend new managerial initiatives, especially as these relate to the work force. Again, TCE concurs but pushes the logic one step further: if and as unanticipated consequences become known they will thereafter be taken into account and mitigated in cost-effective degree.

2.2.3 Law

Contributions to labor law from 1940-1960 included work by the "industrial pluralists" (Archibald Cox, Harry Shulman, and Arthur Goldberg) to develop a more scientific approach to labor law where the types of union organization varied with the attributes of the work force (be

they generic or firm-specific). Legal Realism also continued to critique of mainline law school training but, unable to advance a coherent rival theory, “ran itself into the sand” as an intellectual force (Schlegel, 1979, p.459).

2.3 The 1960s

2.3.1 Economics

As things would play out, two important articles in the 1960s would not only salvage the concept of transaction cost from its enfeebled state but would do so in a way that was embarrassing to the economics profession. Both articles approached the issue by indirection – by accepting the prevailing view that transaction costs were nil and pushing the logic of zero transaction costs to completion.⁵

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, an astonishing result materialized: "Pigou's conclusion (and that of most economists of that era) that some kind of government action (usually the imposition of taxes) was required to restrain those whose actions had harmful effects on others (often termed negative externalities)" was incorrect (Coase, 1992, p. 717).⁶ That is because if transaction costs are 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, p. 717). Arrow's 1969 examination of "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-market Allocation" likewise revealed a need to make a place for positive transaction costs, both with respect to market failures and in

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

But while 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 (which is the Fischer critique (see note 3). 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.

2.3.2 Organization theory

Of special importance to TCE was the reprint in 1962 of Robert Michels' book on Political Parties (first published in 1915), which book is subtitled "A Sociological Study of the Oligarchical Tendencies of Modern Democracy." Its most memorable quote comes in the last chapter where Michels advances the Iron Law of Oligarchy: "It is organization which gives rise 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). But that was not the last word. Once

alerted to the oligarchical propensities of organization, the constructive lesson to be gleaned is this: “Nothing but a serene and frank examination of the oligarchical dangers of ... [organization] will enable us to minimize these dangers” (Michels, 1962, p. 370). In effect, Michels prefigured the concept of credible contracting.

2.3.3 Law⁷

The technical versus purposive distinction made by Llewellyn (1931) was elaborated by Clyde Summers, who distinguished between “black letter law” on the one hand and a more circumstantial approach to law on the other. “The epitome of abstraction is the Restatement, which illustrates its black letter rules by transactions suspended in midair, creating the illusion that contract rules can be stated without reference to surrounding circumstances and are therefore generally applicable to all contractual transactions” (Summers, 1969, p. 566). Such a conception does not and cannot provide a “framework for integrating rules and principles applicable to all contractual transactions” (Summers, 1969, p. 566). A broader conception of contract, with emphasis on the affirmative purposes of the law and effective governance relations, is needed if such is to be realized.

Another interesting legal contribution is Stewart Macaulay's empirical study of contract. Macaulay observed that contract execution is normally a much more informal and cooperative venture than legalistic approaches to contracting would suggest. As one businessman in his study reported, “you can settle any dispute if you keep the lawyers and accountants out of it. They just do not understand the give-and-take needed in business” (1963, p. 1661). More generally, Macaulay's study of contract-in-practice supports the view that contractual disputes and ambiguities are often settled by private

ordering rather than by appeal to the courts – which is in sharp contrast with the neoclassical assumptions in both law and economics.

What follows is my interpretation of the approach that I adopted, as influenced by my interdisciplinary training at Carnegie, my teaching industrial organization, my experience with the Antitrust Division, and my propensity to ask the question, “What is going on here?” as the project unfolded and one thing led to another.

3. Foundations

3.1 How are human actors described?

Herbert Simon was emphatic: “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; emphasis added). That is very different from common practice in economics where the attributes of human actors are introduced after the fact, in support of rather than to inform the choice of research methods -- at the sacrifice, often, of plausibility.

To be sure, behavioral assumptions are not an issue for simple market exchange where competition is reliably efficacious. But while that applies to a great deal of economic activity, it does not apply to all. Indeed, as discussed in Section 4, a large numbers (competitive) bidding condition at the outset sometimes gives way to a small numbers supply relation thereafter. But for the cognitive and self-interest seeking behavioral assumptions upon which TCE rests, the appearance and relevance of such a transformation would have fallen under the radar. Naming the behavioral assumptions at the outset and thereafter working up the contractual ramifications is thus very different

from putting analytically tractable models first, the behavioral assumptions of which are inferred later – as after-thought or, possibly, as an exercise for the reader.

Be that as it may, TCE takes Simon's advice seriously and describes the cognitive and self-interest attributes of human actors as bounded rationality and opportunism, respectively. All complex contracts are incomplete by reason of the first of these and strategic behavior that had been suppressed for 100 years (Makowski and Ostroy, 2001, pp. 481-483, 490-491) makes its appearance when joined by the second.⁸

3.2 What is the lens through which economic organization is examined?

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

TCE employs a contractual approach to which the lens of contract applies. It furthermore distinguishes between the lens of contract/ex ante incentive alignment and the lens of contract/ex post governance. Whereas most contractual approaches to economic organization work out of an ex ante incentive alignment setup (agency theory, mechanism design, modern property rights theory), TCE adopts the ex post governance perspective (with emphasis on maladaptation during the contract implementation interval).

3.3 What is the main problem of economic organization?

The main problem can be variously described. TCE takes it to be adaptation, where provision is made for adaptations of both autonomous and coordinated kinds. Friedrich Hayek (1945) described the autonomous adaptation of economic agents to changes in relative prices as the “marvel of the market.” Chester Barnard, by contrast, was interested in coordinated adaptation of a “conscious, deliberate, purposeful“ kind (1938, p.4) as delivered by internal organization, which can be interpreted as the “marvel of hierarchy.” Adaptations of both kinds, but especially coordinated adaptations, were slighted by mainline price theory.

3.4 What are the core disciplines upon which TCE rests?

TCE is an interdisciplinary project in which economics and organization theory were featured from the outset and aspects of the law (especially contract law) were introduced later. As among these three, economics is the first among equals. But for my training at Carnegie, however, I do not think that I would have been alert to the ideas that organization is both important and is susceptible to analysis (March and Simon, 1958) – in which event I have grave doubts that I would have undertaken in 1970 to revisit vertical integration from a combined economics and organizational perspective.

My experience in teaching Industrial Organization and with antitrust enforcement were both relevant in this respect. Thus the prevailing IO view was that, except as contracting practices and organizational structures had a physical or technical basis, nonstandard and unfamiliar forms of contract and organization were deeply problematic and presumptively anticompetitive. My Carnegie training influenced both my classroom teaching and my interpretation of anticompetitive behavior when I was Special Economic Assistant to the Head of the Antitrust Division of the U.S. Department of Justice.

Exceptional as the leadership of the Antitrust Division was,⁹ the propensity to ascribe anticompetitive purpose to contractual and organizational “irregularities” was palpable. My efforts to make the case that affirmative purposes could be and often were served by nonstandard contractual and organizational forms collided with the prevailing “inhospitality tradition” and did not succeed. I therefore decided to make the study of non-standard and unfamiliar contractual practices and organizational structures priority topics when I resumed teaching.

3.5 What was the implicit methodology?

Although methodology is not a subject that I have been expressly concerned with until recently (Williamson, 2009a, 2009b), the implicit methodology that informs TCE can be interpreted as a splice of Robert Solow (2001) and Milton Friedman (1953) with aspects of Tjalling Koopmans (1957), Nicholas Georgescu-Roegen (1971), and James Buchanan (2001) blended in.

Solow's very brief (two page) statement of pragmatic methodology names three key features: keep it simple, get it right, and make it plausible. To this Friedman, Koopmans, and Georgescu-Roegen add a fourth: derive refutable implications and submit these to empirical testing. TCE implicitly subscribed to all four from the outset, where the fourth would serve as the cutting edge.

3.5.1 Why keep it simple?

Solow observes with reference to the simplicity precept that “the very complexity of real life ... [is what] makes models so necessary” (2001, p. 411). Keeping it simple requires the student of complexity to prioritize, which is not merely important but is also practicable if “Most phenomena are driven by a very few central forces. What a good

theory does is to simplify, it pulls out the central forces and gets rid of the rest“ (Friedman, in Snowdon and Vane 1997, p. 196). Central features and key regularities are uncovered by the application of a focused lens – or, in the spirit of pluralism, focused lenses.¹⁰

3.5.2 What does getting it right entail?

Getting it right entails working up the logic by “translating economic concepts into accurate mathematics (or diagrams, or words) and making sure that further logical operations are correctly performed and verified“ (Solow, 2001, p. 112).¹¹

Note that full formalism (mathematics) is not required from the outset. The statement of Newell quoted earlier that theories are nurtured and are built up is consistent with the TCE concept of a “natural progression“ – from informal to pre-formal to semi-formal and fully formal analysis. Thomas Kuhn concurs: the “early versions of most new paradigms are crude“ (Kuhn, 1970, p. 156), often heuristic models (Kuhn, 1970, p. 184). Words, diagrams, and mathematics correspond to pre-formal, semi-formal, and fully formal analysis, respectively.

3.5.3 Why make it plausible?

Logical consistency is a virtue, but “not everything that is logically consistent is credulous“ (Kreps, 1999, p. 125). Indeed, fanciful constructions that lose contact with the phenomena are suspect – all the more so if more veridical models yield refutable implications that are more congruent with the data.¹²

3.6 Why the insistence on prediction and empirical testing?

Solow omits this fourth precept, but Friedman and Georgescu-Roegen regard it as central. Friedman contends that the performance of a theory “is to be judged by the precision, scope, and conformity with experience of the predictions it yields“ (1953, p. 4). Georgescu-Roegen has a felicitous way of putting it: “The purpose of science in general is not prediction, but knowledge for its own sake,“ yet prediction is “the touchstone of scientific knowledge“ (1971, p. 37). Later, if not sooner, all would-be theories are held to the same standard: stand up and be counted by making predictions and submitting these to empirical testing. This most demanding of the four precepts is crucial for sorting the sheep from the goats.

Indeed, most economists know in their bones that theories that are congruent with the data are more influential. Milton Friedman's reflections on a lifetime of work are pertinent: “I believe in every area where I feel that I have had some influence it has occurred less because of the pure analysis than it has because of the empirical evidence that I have been able to organize.”¹³

4. Implementing the TCE Project: Basics

The development of TCE in the 1970s can be thought of as working at the pre-formal stage in the natural progression where many of the key ideas for the theory are uncovered and a logic of organization begins to take shape. What I take to be the key moves are described here.

4.1 What is the paradigmatic transaction?

Jon Elster counsels social scientists to focus on mechanisms rather than general theories (1994, p. 75). That is very much in the spirit of TCE where puzzling conditions are examined by working up the microanalytics.

A key move for TCE was to take the intermediate product market transaction (the make-or-buy), rather than the employment relation, to be focal. Recall in this connection that Coase (1988) conceded that his early emphasis on the employment was misleading and Simon (1952) described the employment contract as atypical.

My reason for examining the intermediate product market transaction in my 1971 paper “The Vertical Integration of Production“ was not, however, because I anticipated that this would be a prototype for TCE but rather I regarded it as the obvious way by which to correct the well-intentioned but wrong-headed enforcement of antitrust toward vertical integration and vertical market restrictions. My paper had the simple purpose of examining whether respect for organization was important to a better understanding of vertical market structures. That vertical integration would be the focal transaction for what would become TCE was not contemplated but was a fortuitous choice.

4.2 What lens should be employed?

My examination of the intermediate product market transaction as focal had another fortuitous consequence. Once the make-or-buy decision was posed it became natural to interpret firm and market as alternative modes of governance. As a consequence, the orthodox lens of choice gave way to the lens of contract.

To be sure, the lens of contract was “in the air“ – what with developments in social choice, general equilibrium theory, mechanism design, and critiques of zero transaction cost reasoning in the 1960s. But there were differences. Rather than work out of the more widely used lens of contract/ex ante incentive alignment, firm and market organization were examined through the lens of contract/governance. That had the advantage of bringing the comparative contractual differences between firm and market

organization to the surface. The disadvantage is that the formal modeling of TCE with respect to governance is made more difficult.¹⁴

4.3 How are firm and market organization described?

Rather than describe firm and market with respect to the Resource Allocation Paradigm (Reder, 1999) or ex ante incentive alignment, firm and market are regarded instead as alternative modes of governance, where governance is described in terms of the three principles of conflict, mutuality, and order to which Commons (1934, p. 4) referred. Specifically, ex post governance is the means by which to infuse order, thereby to mitigate conflict and realize mutual gains.

Firms and markets, so described, differ in discrete structural ways – as a result of which each has distinctive differences with respect to the main problem of organization, namely adaptations of autonomous and coordinated kinds, as discussed in 3.3 above. (The operationalization of this concept of firms and markets is discussed further in Section 5.)

4.4 What is the main purpose of the organization?

Economizing on transaction costs is taken to be the main purpose (which is not to say the only purpose) or economic organization. Inasmuch, however, as (1) transactions differ in their adaptive needs (as between autonomous and coordinated adaptation), (2) governance structures differ in the adaptive capabilities, and (3) maladaptation is inefficient and undermines the viability of a commercial enterprise, economizing is central. This is captured, for TCE purposes, in the discriminating alignment hypothesis: transactions, which differ in their adaptive needs, are aligned with governance structures,

which differ in their costs and adaptive competencies, so as to effect an efficient (transaction cost economizing) alignment.

4.5 How did the 1971 paper address the puzzle of which transactions go where and why?

In effect, I adopted Peguy's recommendation that complex phenomena be examined in a “modest, slow, molecular, definitive“ way.¹⁵ I therefore sought the answer by examining the details of transactions and of modes of governance in a comparative contractual way. The object was to ascertain the circumstances where outsourcing (procurement of a good or service in the market) experienced contractual problems that could potentially be relieved by taking the transaction out of the market and organizing it internally. In effect, I assumed that “in the beginning there were markets“ and attention was focused on when markets would give way to hierarchy as (yet to be identified) consequential complications set in.

Taking simple market exchange where “faceless buyers and sellers ... meet ... for an instant to exchange standardized goods at equilibrium prices“ (Ben-Porath, 1980, p. 4) as the starting point, the quest was for ties that bind. What are the factors that are responsible for such ties?

My 1971 paper examines a series of plausible conditions that might warrant the move from market to hierarchy only to ascertain that most of them did not survive microanalytic scrutiny in the Peguy tradition. But there was one exception: the procurement in the market for goods or services that require durable, nonredeployable investments in transaction specific assets under circumstances where (1) the parties are boundedly rational, hence long-term contracts are incomplete, (2) the parties will defect,

by reason of opportunism, from the spirit of cooperation to insist on the letter of the contract for outlier disturbances for which the stakes are great, and (3) adjudication in the courts is costly and cumbersome. Although the move from market to hierarchy incurs costs of its own (in incentive intensity and bureaucratic cost respects), the argument is that the comparative costs eventually shift in the direction of hierarchy as asset specificity builds up.

The Fundamental Transformation – whereby what had been a large numbers bidding competition at the outset is transformed into a small numbers supply relation during contract implementation and at the contract renewal interval by reason of investments in durable, nonredeployable assets – is the culprit. (The contract implementation differences between market and hierarchy are discussed in 5.3.5, below.)

4.6 What common contractual assumptions are disputed by TCE?

4.6.1 Two-way common knowledge

The assumption of two-way common knowledge between buyer and seller is routinely invoked by many students of economic organization as through it were uncontroversial, hence needs no explanation. Granted, the parties know a lot about the transactions in which they are engaged and of the disturbances that influence payoffs. To assume, however, that there are no knowledge differences between them is very strong.

Even, moreover, if both parties do possess common knowledge, this can be obfuscated by denial. If one party declares that the state realization is X and the other declares Y and an impasse develops (because the payoffs differ), how is this to be resolved? Lacking the wisdom of King Solomon or three-way common knowledge to

include costlessly informed courts, how is this accomplished? (And even if they are costlessly informed, problems could still arise if the courts are not honest brokers.)

I conclude that two-way common knowledge is implausible, three-way common knowledge is preposterous, and that both can be undone by strategic behavior.

4.6.2 Costless bargaining

Costless bargaining is another manifestation of zero transaction costs. As such, this is a red flag. What are the preconditions for costless bargaining to be satisfied? If this is merely a matter of analytical convenience, then the refutable implications of the theory become all the more important. What are they? And are the data corroborative?

5. Refinements: Implementing the TCE Project

One thing leads to another. Having worked through the intermediate product market transaction in transaction cost terms, I began to perceive that this comparative contractual approach to vertical integration could also be applied to other commercial phenomena. But I also recognized that unexamined conceptual issues needed to be addressed for which gap filling or explication was warranted. Follow-on work in both breadth and depth respects was thus undertaken, where the latter would entail the explication of core concepts and the introduction of others.

5.1 Depth

Although the issues discussed in this subsection are all examined in relation to the intermediate product market transaction, all also have relevance to the study of complex contracting more generally.

5.1.1 Dimensionalization: transactions

What are the key attributes of transactions?

As a scan of the literature reveals, it is easy to name a unit of analysis. What is difficult and commonly observed in the breach is to identify the critical dimensions with respect to which the unit of analysis differs. Dimensionalizing transactions was featured by TCE from the outset.

The condition of asset specificity (including the various forms that it can take – physical, human, site specificity, dedicated, brand name capital, and temporal), the disturbances to which adaptations are subject, and the frequency of transactions are taken to be the key features – with emphasis on the first two. Asset specificity is the source of bilateral dependency and, as such, would become the “big locomotive“ upon which TCE relied. Disturbances pose the need for adaptation. Both were featured from the outset and were successively elaborated as TCE progressed. The combination of asset specificity and outlier disturbances in incomplete contracts for which the stakes are great posed the need for coordinated adaptation – whereupon markets give way to hierarchies.

5.1.2 Dimensionalization: governance structures

What are the key attributes for describing modes of governance?

The two key attributes of governance structures upon which TCE initially relied are incentive intensity and administrative command and control. Contract law regime differences would be subsequently added.

The basic argument is that modes of governance differ in discrete structural ways (Simon, 1962, 1977) and that each is defined by an internally consistent syndrome of attributes. As between markets and hierarchies, the incentive intensity differences are that markets harness high-powered incentives (each party appropriates its net receipts)

whereas hierarchy employs low-powered incentives (in support of uncontested coordination at the interface). The command and control differences at the interface are that these are negligible for the market mode but strong for hierarchy.

As discussed in 2.1.3 and 2.3.3, the contract law of markets is that of legal rules and this gives way to contract as framework as the continuity of the exchange relation builds up. It was not, however, until I expressly examined markets, hierarchies, and hybrids in my paper on “Comparative Economic Organization: The Analysis of Discrete Structural Alternatives” (1991) that I perceived the need to include differences in contract law regime as a third dimension of governance.

In addition, therefore, to simple market exchange (to which contract as legal rules applied) and supporting ongoing exchange (to which contract as framework applied), I now asked the question: What is the implicit law of contract for unified ownership? My answer was this (Williamson, 1991, p. 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 thus introduced to fill a gap in the theory of governance. The efficacy of all forms of contract law, forbearance included,

varies with the integrity of the institutional environment (nation state) of which it is a part.

Markets and hierarchies are thus polar opposites in incentive intensity, administrative command and control, and contract law respects. The hybrid is a compromise mode on all these attributes. The importance of providing credible contracting supports for the hybrid mode became evident in this connection.

5.1.4 Credible contracting and the Simple Contractual Schema

How does credibility figure in the TCE setup?

The concept of credible commitment goes back to antiquity but took on prominence in Thomas Schelling's famous essay on bargaining (1956) and was subsequently developed in conjunction with barriers to entry (Dixit, 1979, 1982; Eaton and Lipsey, 1981; Schmalensee, 1981) and reputation effects (Kreps and Wilson, 1982; Milgrom and Roberts, 1982). My purposes were different. The TCE needs for credible commitment were perceived to be especially great in conjunction with the hybrid mode of contracting. One of the consequences is that many of the mechanisms that had previously been interpreted as anticompetitive are now perceived to be a means by which to promote more efficient contracting for transactions of the middle kind (Williamson, 1983).¹⁶

An illustration of the mechanisms that would support credibility for hybrid contracts is provided by the 32-year coal supply agreement between the Nevada Power Company and the Northwest Trading Company. That contract reads in part as follows (Williamson, 1991, pp. 1238-129):

... In the event an inequitable condition occurs which adversely affects one Party, it shall then be the joint and equal responsibility of both Parties to act promptly and in good faith to determine the action required to cure or adjust for the inequity and effectively to implement such action. Upon written claim of inequity served by one Party upon the other, the Parties shall act jointly to reach an agreement concerning the claimed inequity within sixty (60) days of the date of such written claim. An adjusted base coal price that differs from market price by more than ten percent (10%) shall constitute a hardship. The Party claiming inequity shall include in its claim such information and data as may be reasonably necessary to substantiate the claim and shall freely and without delay furnish such other information and data as the other Party reasonably may deem relevant and necessary. If the Parties cannot reach agreement within sixty (60) days, the matter shall be submitted to arbitration.

By contrast with a classical contract, this contract (1) contemplates unanticipated disturbances for which adaptation is needed, (2) provides a tolerance zone (of $\pm 10\%$) within which misalignments will be absorbed, (3) requires information disclosure and substantiation if adaptation is proposed, and (4) provides for arbitration in the event voluntary agreement fails.

The forum to which the hybrid contract refers disputes is (initially, at least) that of arbitration rather than the courts. Fuller (1963, pp. 11-12) described the procedural differences between arbitration and litigation:

[T]here are open to the arbitrator ... quick methods of education not open to the courts. An arbitrator will frequently interrupt the examination of witnesses with a

request that the parties educate him to the point where he can understand the testimony being received. This education can proceed informally, with frequent interruptions by the arbitrator, and by informed persons on either side, when a point needs clarification. Sometimes there will be arguments across the table, occasionally even within each of the separate camps. The end result will usually be a clarification that will enable everyone to proceed more intelligently with the case.

Such adaptability notwithstanding, neoclassical contracts are not indefinitely elastic. As disturbances become highly consequential, hybrid contracts give way to hierarchy.

Interpreting credible contracting as a means by which to describe the hybrid mode of economic organization would lead in turn lead to the “simple contractual schema“ (see Figure 1), where price, asset specificity, and contractual safeguards are all determined simultaneously. Taking k to be an index of asset specificity, $k = 0$ denotes the absence of specialized investments, in which case the parties are essentially faceless. Those transactions that use the special purpose technology are ones for which $k > 0$. As discussed earlier, 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 by simple market

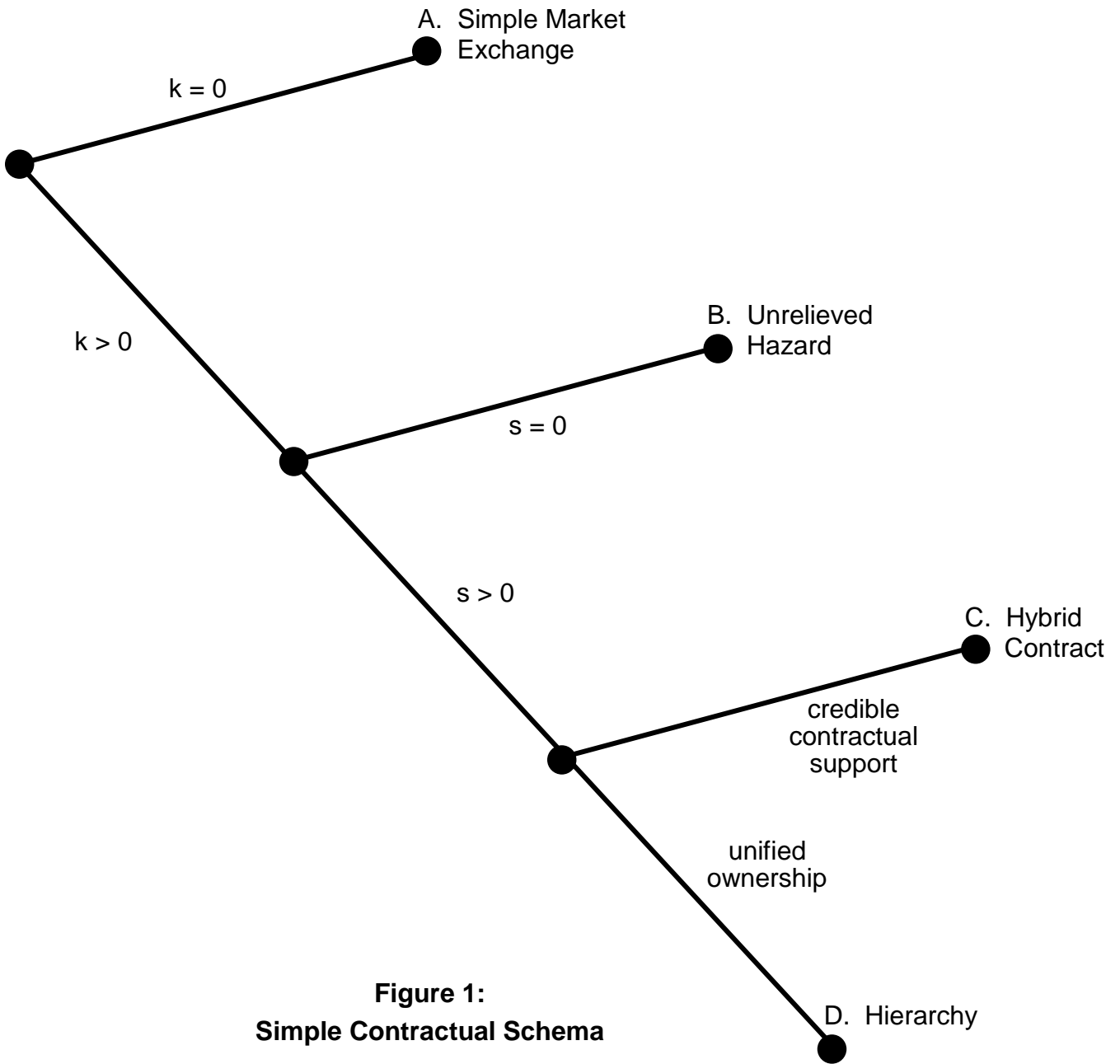


Figure 1:
Simple Contractual Schema

exchange and, in the event of disputes, by court awarded damages. 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, the prospect of costly contractual breakdowns continue in the face of best efforts to craft safeguards at node C, the transaction will 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 pose greater needs for cooperative adaptation.

5.1.5 Scaling up¹⁷

Does the simple model scale up?

The object of a simple model is to capture the essence, thereby to interpret puzzling practices and make predictions that are subjected to empirical testing. But that is not the only relevant test. Simple models can also be "tested" with respect to scaling up. Does repeated application of the basic mechanism out of which the simple model works yield a result that recognizably describes the phenomenon in question?

The test of scaling up is often ignored, possibly out of awareness that scaling up of the model in question is very demanding. Sometimes it is recognized but deferred,¹⁸

possibly in the belief that scaling up can be accomplished easily. Whatever the reason for evading the issue, claims of real world relevance, including public policy relevance, of any candidate theory of the firm that cannot be shown to scale up from toy model status to approximate the phenomenon of interest (usually, the modern corporation) should be met with skepticism.

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

5.1.6 The institutional environment

How does TCE relate to the institutional environment?

The implicit institutional environment to which TCE is usually applied is that of Western capitalist countries. Plainly, however, the Western institutional environment evolved over time and differences among Western capitalist countries warrant attention, which differences are all the more so between capitalist and noncapitalist countries.

The New Institutional Economics makes express provision for both the institutions of governance (the rules of the game) and the governance of contractual relations (the play of the game). The rules of the game are broken down into the informal rules (customs, traditions, morays, religion) and the formal rules (the polity, judiciary, laws of property and contract).¹⁹ Differences between nation states in formal and informal rules respects are pertinent to the advisability of undertaking some complex transactions and to the efficient alignment of transactions and governance structures.

An example of a failure to respect for the rules of the game as these relate to property rights and the hazards of expropriation is Mikhail Gorbachev's advice to American firms to invest quickly, rather than wait: "Those [companies] who are with us now have good prospects of participating in our great country [whereas those who wait] will remain observers for years come – we will see to it."²⁰ That the leadership of the Soviet Union "will see to it" that early and late movers will be rewarded and punished, respectively, reflects conventional carrot-and-stick incentive reasoning. What it misses is that nation states where the leadership can reward friends and punish others pose investment hazards to all – in that what is given can subsequently be taken away by muscular abuses, to include asset seizure. The upshot is that fewer degrees of freedom (rules) can have advantages over more (discretion) if cost effective credible commitments accrue to the former.

Economic reform should thus be informed by efficiency reasoning – to which credible contracting, or its absence, is pertinent.

5.2 Breadth

Does vertical integration serve as a paradigm problem, in that many other commercial (and some noncommercial) transactions can be interpreted as variations on a theme?

As stated at the outset, any issue that arises as or can be reformulated as a contracting problem can be examined to advantage in transaction cost economizing terms. The value added of TCE varies, however, from much too little, depending on the particulars. Five such examples are offered here, after which I briefly discuss uses of TCE by the contiguous social sciences.

5.2.1 Significant reformulations

Both franchise bidding for natural monopoly and exchange agreements are examples of how applications of the lens of contract/governance led to significant reformulations.

The main purposes of my paper on franchise bidding for natural monopoly (Williamson, 1976) were to ask and answer three questions: Does the efficacy of franchise bidding as described by Demsetz (1968) in his examination of automobile license plates extend to the much more complex case of CATV as claimed by Posner (1972)? Where do the main conceptual difficulties reside? And what does the evidence support?

Given my awareness that complex contracts are much more problematic than simple contracts, I had doubts from the outset that Posner's claims about the efficacy of franchise bidding for CATV circa 1970 would withstand scrutiny. For one thing, I was aware from my late 1960s experience on Mayor Lindsay's CATV Task Force on CATV that numerous economic and political complications were posed when making CATV

awards. For another, Posner's position that "To expound the details of particular regulations and proposals ... would serve only to obscure the basic issues" (1972, p. 98), ran contrary to my work on TCE in the 1970s on the importance of examining the details. Also, Posner made no provision for differences between the easy redeployability of assets used for the production of automobile license plates and the complications posed by CATV. Indeed, Posner's sanguine views on the efficiency of franchise bidding for natural monopoly in CATV differed from mine in all of the following respects: (1) the costs of ascertaining and aggregating consumer preferences through direct solicitation; (2) the use of scalar bidding; (3) the degree to which technology is well developed; (4) demand uncertainty; (5) the degree to which incumbent suppliers acquire idiosyncratic skills; (6) the extent to which specialized, long-lived equipment is involved; and (7) the susceptibility of the political process to opportunistic representations and the differential proclivity, among governance modes, to be politicized.

I concluded that both in the abstract and in the particulars -- as borne out by reports by the Cable Television Information Center²¹ and by my focused case study of the experience with franchise bidding for CATV in Oakland, California in the late 1960s and early 1970s²²--that Posner was unduly sanguine.

Another example of how TCE can inform the understanding of complex contracting is that of exchange agreements. My interest in this originated with discussions that I had in the early 1970s with petroleum engineers, lawyers, and managers who took the position that an exchange agreement between company A, which had a surplus in area I and a deficit in Area II, and company B, which had a deficit in area

I and a surplus in area II, was efficient because it saved on transportation expense. My question as to why these savings could not be realized by reporting surpluses and deficits to a central market was met with frowns and silence.

Might there be transaction cost reasons to engage in exchange? That seemed plausible, but both lacked a theoretical basis and microanalytic underpinnings until the 1980s when I read the study of the Canadian petroleum industry done by the Canadian antitrust authorities at the same time as I was working on credible commitments. The study included documents from the files of the petroleum firms that were interpreted by the Canadian authorities as having anticompetitive purpose and intent, an example of such being a the following quotation from a petroleum firm's files: "We do believe the oil industry generally, although grudgingly, will allow a participant who has paid his ante, to play the game; the ante in this game being the capital for refining, distributing, and selling products."²³ Examined in conventional market power (lens of choice) terms, this language certainly sounds anticompetitive. Examined through the lens of contract/governance, however, another interpretation suggests itself: established firms are willing to sell to firms that have made investments in the industry that signal participation of an ongoing kind but not to fly-by-night firms that are doing deals of an episodic kind. (Also, constructing capacity that exceeded a firm's own current needs could be efficient if economies of scale are realized and trades with a rival firm in a different area can be arranged, especially if the two firms experience positively correlated demand disturbances.)

5.2.2 Complementary reformulations

Two complementary reformulations are discussed here: TCE interpretations of the employment relation and of corporate finance transactions.

TCE does not address the employment relation with reference to the zone of acceptance – as in Barnard (1938) and Simon (1952) – where the zone of acceptance describes the range of activities over which the worker agrees to accept authority in return for the agreed upon terms of compensation. Rather, TCE examines when and why employment relations are supported by differing forms of collective organization – which is more in the tradition of the “industrial pluralists” (Archibald Cox, Harry Schulman, Arthur Goldberg) and the internal labor market literature (Doeringer and Piore, 1971).

Granting that there is a role for collective organization (often in the form of unions), the question is when does this move beyond simple negotiations over wages and benefits to include firm-specific productivity enhancement through human asset development, special dispute settlement mechanisms, and special mechanisms to support adaptation. Also, how do the potential hazards of collective organization figure in? The upshot is that cost-effective contracting with unions varies from simple to complex – many along the lines of the TCE setup (Williamson, 1985, ch. 10). As between the efficiency reasoning of the industrial pluralists and Katherine Stone's (1981) power reasoning, TCE comes down in favor of the former (Williamson, 1985, pp. 250-252).

Can TCE also inform the uses of debt and equity? Finance being a highly specialized field, maybe not. If, however, debt and equity are interpreted not merely as financial instruments but also as alternative modes of governance, and if investments by firms vary in their degree of redeployability, possibly there is something there.

Upon describing debt as a rules governed mode of finance whereas equity is much more discretionary and assuming that investments that vary in asset specificity are properly supported by different types of governance, an “efficient alignment“ interpretation of transactions (investments) with governance structures (rules versus discretion) suggests itself. Specifically, the use of TCE reasoning in my article on “Corporate Finance and Corporate Governance“ (1988) predicts that generic investments will (or should) be debt financed whereas firm specific assets should be financed by equity. I also discuss applications to leasing (for assets on wheels) and the efficiency benefits of leveraged buyouts (to correct an inefficient choice of debt and equity where debt is underused by the incumbent management because it is a more demanding mode of governance).

To be sure, the TCE rationale for efficient choice of financial instrument is only one of several reasons to take exception with the Modigliani Miller theorem. But the basic point, here as elsewhere, is this: efficient governance varies systematically with the attributes of transactions.

5.2.3 Servicing the periphery

If TCE really informs any issue that arises as or can be reformulated in contractual terms, does it also apply to marriage?

As I discuss in 5.3 in conjunction with trust, calculativeness can get in the way of highly personal relations, of which marriage is one. TCE being a calculative approach to contracting, marriage is treacherous ground. Consider, however, a variant of marriage that I will refer to as “career marriages,“ of which two kinds are distinguished. One is a marriage of convenience where neither party asks the other to make career sacrifices on

his/her behalf. If, therefore, one party gets a good offer to move to a different location, he/she moves on and the two parties “split“ with no hard feelings.

The other type of career marriage is one where one party (traditionally the husband) asks the other party to make career sacrifices on his/her behalf. This poses risks for the party who agrees to make sacrifices that his/her career will be compromised. Those risks could be mitigated, however, if the party making the request created deterrents to divorce or unseemly behavior (e.g., by setting up a substantial divorce fund). But then again he/she could refuse. Refusal implies that there is no safeguard ($s = 0$) while risk mitigation implies that $s > 0$.

TCE interprets the $s = 0$ condition as that which is memorialized by Carroll Channing: “diamonds are a girl's best friend.” More generally, $s = 0$ is a marriage of living in the moment (travel, nights on the town, fancy wardrobes, jewelry) because there may be no tomorrow. The $s > 0$ condition will be less hurried and less hurried by comparison.

Fanciful? Sort of. Yet it is a way to get TCE across to people who roll their eyes about vertical integration.

5.2.4 Applications to the contiguous social sciences

Does TCE have interdisciplinary applications?

The answer here is decidedly yes, especially with respect to the empirical TCE literature. In addition to applications within applied fields in economics, TCE also has applications within the business schools (in strategy, organization theory, marketing, finance, operations management, and accounting) and to the contiguous social sciences (especially law, sociology, and political science).²⁴

Of these various applications I briefly digress on marketing, where the uses of TCE took hold early and continue to grow – sometimes with reservations but mainly in supportive ways and, in any event, were inspired by TCE (John and Reve, 2010).²⁵ Interestingly, parts of the empirical marketing literature can also be interpreted as support for agency theory, which has been important as a theory in the economics of organization literature but for which empirical tests of which have been slight. Bengt Holmstrom and Paul Milgrom (1991, 1994), for example, interpret Erin Anderson's 1982 empirical research as providing support for agency theory, to which I agree. It is nonetheless noteworthy that this empirical research had TCE origins. So the question is why it did not originate with agency theory?

More generally, TCE has been an empirical success story: “Despite what almost 30 years ago may have appeared to be insurmountable obstacles to acquiring the relevant data [which are often primary data of a microanalytic kind], today transaction cost economics stands on a remarkably broad empirical foundation” (Geyskens, Steenkamp and Kumar 2006). As Michael Whinston puts it: TCE has been “one of the great [empirical] success stories in industrial organization over the past [30] years” (Whinston, 2001, p. 185).

LaFontaine and Margaret Slade concur (2007, p. 658):

The weight of the evident is overwhelming. Indeed, virtually all predictions from transaction cost analysis appear to be borne out by the data. In particular, when the relationship that is assessed involves backward integration between a manufacturer and her suppliers, there are almost no statistically significant results that contradict TCE [transaction costs] predictions.

This is not to say that more work is unneeded. For example, most empirical work on TCE is a reduced form construction. Econometric refinements, as in Scott Masten, James Meehan, and Edward Snyder (1991), are demanding and should be instructive (but probably will not undo the broad base of empirical support upon which TCE rests (Macher and Richman, 2008)).

5.3 New concepts/unmet needs

Implementing TCE is an interesting learning experience as one thing leads to another. My purpose here is to describe some of the perceived needs to introduce new concepts and to recognize, if not resolve, unmet needs. I begin with the former and then briefly describe unmet needs.

5.3.1 Limits to firm size

Why can't a large firm do everything that a collection of small firms can do and more?

The firm size issue has a long history (Knight, 1921; Coase, 1937) and was addressed more recently by Tracy Lewis, who argued that an established firm can always realize greater value because it can “use the input exactly as the new entrant would have used it ... [and can furthermore] improve on this by coordinating production from his new and existing inputs“ (Lewis, 1983, p. 1092). Transaction cost economics takes exception with this argument by identifying and examining the efficacy of the two mechanisms on which Lewis' formulation implicitly relies: replication and selective intervention.

Thus suppose that a buyer acquires a supplier with the understanding that (1) the supplier will operate in the same autonomous way post-acquisition as in the pre-

acquisition status (by replication) except as (2) the buyer intervenes selectively when expected net gains can be achieved through coordinated adaptations. The combined firm can then never do worse (by replication) and will sometimes do better (by selective intervention). Therefore, more integration is always better than less, and repeating this logic implies that everything will be organized in one large firm.

No surprise: the action resides in the microanalytics, which are tedious but revealing (Williamson, 1985, pp. 133-142). Very briefly, none of the four implicit “promises“ upon which replication and selective intervention rely is self-enforcing – those assumptions being: (1) the buyer as owner promises the acquired supplier its net receipts in all state realizations – thereby preserving strong incentives; (2) the supplier promises to use the supply stage assets that the buyer now owns with “due care”; (3) the buyer promises to exercise authority (fiat) only when expected net gains can be ascribed to selective intervention; and (4) the buyer promises to reveal and divide the net gains from selective intervention as stipulated in the original agreement.

The problem is that none of these promises is self-enforcing. To the contrary, in the absence of perfect information (to include a costless arbiter), each condition will be compromised.

5.3.2 Market failure remedies

When do market failures warrant government corrective action?

Much of the extensive economics literature on market failure suggests that the answer to that question is that government intervention should be frequent. Often, however, that is due to a one-sided treatment of market failure.

The asymmetry is this: whereas it is uncontested that markets experience failures, a similar orientation is not apply to the public sector. To the contrary, “normative public policy analysis began by assuming that ... policy was made by an omnipotent, omniscient, and benevolent dictator“ (Dixit, 1996, p. 8). Omniscience wipes out bounds on rationality, benevolence eliminates opportunism, and omnipotence eliminates implementation obstacles. Such asymmetry supports a propensity for public policy to intervene whenever a market failure is observed – which harkens back to Coase (1960, 1964).

TCE proposes to remedy this condition by introducing the Remediable Criterion, namely: an extant mode of organization for which no superior feasible alternative can be described and implemented with expected net gains is presumed to be efficient. Hypothetical ideals thus give way to descriptions of a feasible alternatives; implementation obstacles, economic, political, and otherwise, need to be factored in under the net gain criterion; yet the presumption that current practice should be continued if no net gains can be displayed, is rebuttable upon a showing that unfair or otherwise unacceptable obstacles are the incumbent. Politics thus trumps economics in political regimes that are not seriously corrupted (are above threshold) but needs to make that explicit.

The Remediable Criterion has the salutary effect of requiring those who propose reforms to do their homework. There being an endless number of worthy causes, each needs to be examined in a hard-headed way. For TCE purposes, an examination of the relevant microanalytics is basic.

5.3.3 Trust

What is trust?

Trust is a good word, but does it promote or obscure our understanding of complex commercial phenomena? Evidently many social scientists believe that trust is under-used yet has broad reach and have begun to describe situations of trust as “a subclass of those involving risk. They are situations in which the risk one takes depends on the performance of another actor“ (Coleman, 1990, p. 91). According to this formulation, trust is warranted when the expected gain from placing oneself at risk to another is positive, but not otherwise. Indeed, the decision to accept such a risk is taken to imply trust (Coleman, 1990, p. 105). So construed, previously fuzzy conceptions of trust are purportedly clarified and made more operational when trust is treated as a subclass of calculated risk.

I contend that conflating trust and calculated risk frequently leads not to clarification but to obfuscation, as with James Coleman's ex post interpretations of good outcomes as ones of trust and bad outcomes as misplaced trust (Williamson, 1985, pp. 257-258; 262-267).²⁶

My suggestion is that trust be reserved for close personal relationships that would be devalued by calculativeness and that commercial relations and personal relationships of convenience be treated in a calculative way (for which risk and the calculation of expected net gains are appropriate) (Williamson, 1985, pp. 272-273). (Similar reasoning applies to other user-friendly words, of which fairness is one and reciprocal altruism (which is an oxymoron) is another). User-friendly social science that obfuscates comes at a high cost.

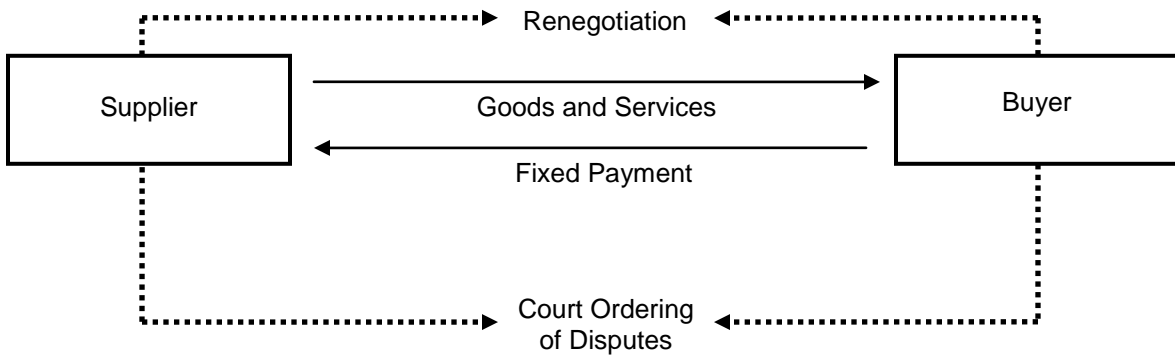
5.3.4 Interface management

Can diagrammatic support be provided for the interface management differences between markets from hierarchies?

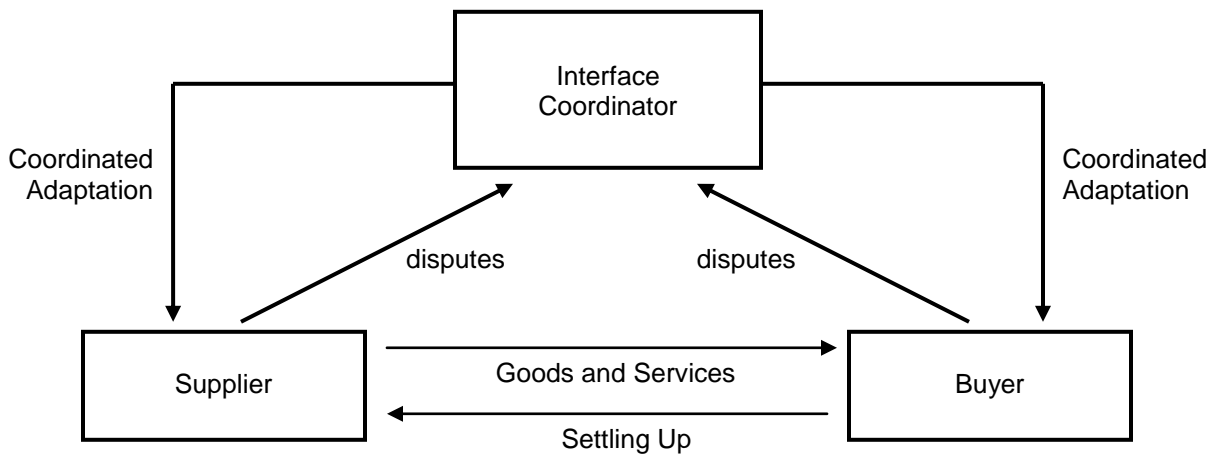
Figure 2 provides a heuristic interpretation of the interface differences between markets and hierarchies for implementing intermediate product market transactions.²⁸ Incentive intensity, adaptation, and dispute settlement differences are crucial.

Panel 1 describes simple market exchange where goods or services are delivered to a buyer in exchange for a stipulated payment, each party appropriates its net receipts, adaptations to state realizations for which the contract is silent or defective are accomplished by renegotiation, and irreconcilable disputes are dealt with by the courts which apply the appropriate legal rules to award money damages.²⁹ The upshot is that high-powered incentives, adaptation by renegotiation, and legalistic dispute settlement are all operative in Panel 1.

By contrast, Panel 2 (hierarchy) introduces a new actor, the interface coordinator, to which each stage reports and receives administrative direction and control. Coordinated adaptation (the need for which could be perceived by buyer or supplier stages or undertaken at the initiative of the interface coordinator) is made with reference to expected net gains and is done in a timely way (without adversarial bargaining and without review by the courts) on the decision of the interface coordinator. Disputes are also mediated by the interface coordinator on the merits (private ordering as supported by forbearance law). And the payments to the supplier and buyer stages are made from a common treasury with the object of making them whole but with little else, the effect of which is to provide low-powered incentives and uncontested compliance by each stage.



Panel 1:
(Simple) Market Mediated Exchange



Panel 2:
(Continuity Preserving) Hierarchy Mediated Exchange

Figure 2. Interface Mediation

Markets and hierarchies, so described, are polar opposites in incentive intensity, coordinated adaptation, and dispute settlement respects. The introduction of the interface coordinator for $k > 0$ transactions is truly consequential.

5.3.5 Burdens of bureaucracy

What are the burdens of bureaucracy and what do we do about them?

Bureaucracy is one of the most important but poorly understood conditions in the social sciences. One common mistake is to condemn bureaucracy because it falls short of a hypothetical ideal. Because, however, all feasible modes of organization are flawed, the burdens of bureaucracy need to be examined comparatively and evaluated with reference to a net gain criterion. TCE has addressed some of these issues, but in a piecemeal way.

Because the putative purposes of bureaucracy are usually commendable, because the downside of bureaucracy is amorphous and resides in the future, and because politicians often have a short horizons, the burdens of bureaucracy often become someone else's problem by pushing them off to another day -- an example of which is Joe Nocera's visit to the recently organized Consumer Financial Protection Bureau, which he described to his mentor as an "inspiring day." His mentor responded with the observation that "The beautiful thing about a new agency is that everyone is very driven to accomplish the mission"-- to which, however, he added "As they mature that's when people become more concerned with self protection, and maneuvering for the next promotion." Nocera's response was "True enough, but a problem for another day" (2012, p. A23).

This last has it exactly wrong. Predictable consequences should be factored in at the outset. Failure or refusal to do so shifts the bureaucratic burden to the successor generations – for which the task of correcting badly performing bureaus is not only costly but doubtful because such efforts face great resistance from those who have a stake. Indeed, terminating bloated bureaus is well-nigh impossible because the beneficiaries of termination are diffuse and difficult to organize in comparison with the bureaucrats and politicians and beneficiaries of the bureau who are fewer in number and individually have big stakes.

Note, however, that such arguments operate at a very high level of generality. Aside from the common concern that the burdens of bureaucracy build up over time and are sheltered by the difficulties of reversing regulatory decisions, what are the lessons? In the spirit of pluralism, this is surely a case where different lenses should be brought to bear in the hope that someone will “get it right.”

The TCE approach is to move beyond overarching critiques of bureaucracy to examine the particulars. A basic question that arises in this connection is this: How are the “transactions” for which regulatory solutions are proposed described? I suggest in this connection that administrative transactions (e.g., foreign policy) and regulatory transactions (e.g., pollution) are sufficiently different to warrant separate treatment.

My focus is on the regulatory arena. I conjecture in this connection that regulatory transactions are more variable than commercial transactions, as a consequence of which remedies for regulatory breakdowns are (probably) also more variable.

Regulatory breakdowns can take several forms, including capture (Bernstein, 1955), pre-capture (Stigler, 1971), being overmatched in relation to the expertise of the

industry (e.g., the SEC in relation to financial industry), the propensity for over-reaching (which is common to all in variable degree), and the hazard of false confidence among members of the public who wrongly ignore hazards to which they are subject because they believe that they are safely protected by regulatory oversight and control.

The challenge is to uncover the particulars that are associated with each class of breakdown and ask the following questions of proposed regulatory initiatives: What implementation problems are in prospect? What are the ex ante regulatory design ramifications? And given best efforts to deal with these issues, does the proposed regulation satisfy the Remediableness Criterion? (Also, additional types of breakdown not described above should be included.³⁰)

5.4 Other Issues and Unmet Needs

5.4.1 Other Issues

What issues for which transaction cost economics has made provision are in a seriously underdeveloped state of analysis?

Several that I regard as important but are seriously underdeveloped are (1) the importance of atmosphere as it relates to textbook bromides, of which the uses of marginal cost pricing is one and the advice never to leave any money on the table is another, (2) the need to breathe operational content into the concept of power, (3) making provision for disequilibrium contracting, and (4) the putative benefits of pay-for-performance with respect to executive compensation.

The missing concept with respect to bromides is not that they are unhelpful for classroom purposes but that they need to be put in perspective in conjunction with

applications, lest students are persuaded by their teachers to take them literally. With respect to marginal cost pricing the problem is partly one of measurement obstacles aided and abetted by unanticipated consequences (purposeful biases and distortions) (Williamson, 2005, pp. 26-27).

The bromide never to leave money on the table was contested by an investment banker who informed me that one of the most important negotiating lessons that he had learned was “always to leave money on the table.” The problem with the maxim never to leave money on the table is that meticulous observance by one party invites responses in kind – with the result that what could have been a cooperative contractual relationship becomes much more adversarial as what could have been give-and-take results instead results in contentious ex ante bargaining and ex post disputes.

What distinguishes the advice to “leave no money on the table” and “always leave money on the table” is not that the former is hardheaded and the latter is soft. Rather, in circumstances where cooperation during contract implementation is perceived to be important, the former is myopic and the latter is farsighted.³¹

Disequilibrium contracting is an issue that I briefly raise in conjunction with hybrids yet remains underdeveloped. Implicitly the standard TCE assumption is that choice among alternative modes of contract is done in an unhurried way. In the context of innovative startups or crises, however, the possibility that what would be regarded as efficient modes of governance under equilibrium conditions will often give way to the urgency of real time responsiveness (Williamson, 1991, pp. 292-293).

Similarly, the contention that “lack of strong pay-for-performance incentives for CEOs” (Jensen and Murphy, 1990, p. 262) needs to be examined with respect to the ease

with which boards of directors can be induced to approve the compensation recommendations of executives (Bebchuk and Fried, 2003, p. 73; Williamson, 2005, pp. 31-33). This too awaits further study.

More generally, the hazards of making uninformed choices for failure to give prominent place to the attributes of human actors (in both bounded rationality and opportunism respects) are chronic concerns of public policy toward business.

5.4.2 Unmet needs

What are the unmet needs?

Unmet needs include full formalism, “implementation economics,” and a broader, deeper treatment of bureaucracy.

The full formalism of TCE has been a work-in-progress since the famous paper by Sanford Grossman and Oliver Hart (1986). Steven Tadelis and his collaborators have been especially important in moving the formal analysis of TCE ahead.³¹ This is outside my bailiwick but I raise again the possibility that a “new mathematics” (to which Edward Prescott made reference) is needed.

The need for “implementation economics” was first called to my attention by a financial engineering colleague who suggested that a productive collaboration between financial engineering and TCE was possible. I objected and stated that most users of TCE lacked the technical background to design new financial instruments. My colleague agreed that financial engineers did not need assistance in designing financial instruments but observed that financial engineering was deficient in implementation respects.

Given the emphasis in TCE of ex post contract implementation, and considering the unforeseen problems with derivatives that contributed to the recent financial crisis, I

could see where his concern with implementation was coming from. Given my efforts to understand the purpose served by non-standard contracting, moreover, I could see that the neglect of implementation by imaginative designers might well be the source of avoidable error.

The “problem” was and is that acquiring the requisite knowledge of the relevant microanalytics, which is where the TCE action is concentrated, is very demanding. Inasmuch, however, as the stakes are great, intensive efforts to uncover the prospective “unanticipated consequences” by TCE specialists could well be warranted. Ascertaining whether such is promising or not would benefit by an examination of the recent crisis (with emphasis on the details and real time urgency of the breakdowns), an ex post interpretation of what is going on here in TCE terms, and an assessment of the value added.

Finally, in addition to the burdens of bureaucracy discussed above, the study of bureaucracy is also relevant to an understanding of comparative economic systems. Albeit dated, I call attention to Oskar Lange's treatment of socialism (1938) as an example of what would become standard practice in economics for the next 50 years: concede that bureaucracy was important to an understanding of socialism, yet refuse to address bureaucratic consequences in a systematic way.

Thus Lange's influential treatment of socialism concedes that “the real danger of socialism is that of a bureaucratization of economic life, and not the impossibility of coping with the problem of allocation of resources” (1938, pp. 109-110; emphasis in original). Lange nevertheless declined to examine the importance of bureaucracy because such a discussion “belongs to the field of sociology rather than economic theory

and must therefore be dispensed with here” (1938, p. 109). He furthermore stated that “the same, or even greater, danger” is posed by monopolistic capitalism (1938, p. 110) – although the basis for this is not revealed.

Be that as it may, rather than undertake a comparative study of bureaucracy, the study of comparative economic systems mainly set the burdens of bureaucracy aside, evidently in the conviction that “there is nothing wrong with the pure theory of socialism“ (Schumpeter, 1942, p. 172). (Subsequent theoretical work on activity analysis by Koopmans and Kantorovich would likewise ignore bureaucracy.)

Admitting to the burdens of bureaucracy would have to await the breakdown of the USSR, yet even now a comparative assessment of the burdens of bureaucracy as between capitalism and socialism (and variants within each) remains largely unexamined by economists.

7. Conclusions

(((to come)))

Footnotes

1. As David Kreps would observe, this could be viewed as a radical move in that previously “The firm is like individual agents in textbook economics. ... Agents have utility functions; firms have a profit motive; agents have consumption sets; firms have production possibility sets. But in transaction cost economics, firms are more like markets – both are arenas within which individuals can transact” (1990, p. 96). Treating firm and market as alternative modes of governance was a key TCE move.
2. For a discussion of research orientations in the sciences, see Roy D'Andrade (1986). Robert Solow takes the position that “there is a lot to be said in favor of staring at the piece of reality you are studying and asking, just what is going on here? Economists who are enamored of the physics style seem to bypass that stage” (Solow, 1997, p. 57; emphasis added). The theoretical physicist John Bardeen in his work on the transistor nevertheless reports that they “did a lot of experiments to try to find out just what was going on here” (Hoddeson and Daitch, 2002, p. 138; emphasis added), thereby to better inform their theoretical work. Evidently the difference between the two polar research traditions should not be overdone yet is nevertheless instructive.
3. As Stanley Fischer put it, “Transaction costs have a well-deserved bad name ... [because] there is a suspicion that almost anything can be rationalized by invoking suitably specified transaction costs” (1977, p. 322).

4. Of the 15 research faculty at GSIA when I arrived as a student in the PhD program in 1960, four would go on to receive the Nobel Memorial Prize in Economics for work that they did at GSIA— namely, Herbert Simon, Franco Modigliani, Merton Miller, and Robert Lucas (although Lucas did not arrive until 1963). Also, four graduate students in the GSIA program in the 1960s would also receive Nobel Prizes in Economics.
5. The next two paragraphs are based on Williamson (2010).
6. Even the Chicago School, which had grave reservations with overreaching externality arguments, was resistant to Coase's claims that externalities vanished under zero transaction cost conditions. For a discussion of Coase vs. Chicago, see Edmund Kitch (1983, pp. 220-221).
7. The next two paragraphs are from Williamson (1985, p. 10).
8. Note that it is the combination of both, rather than either by itself, that is needed. Absent bounded rationality, all complex contracts will be complete. Absent opportunism, contract as promise will suffice. Serious contracting problems await if, given both, a condition of bilateral dependency develops.
9. Donald Turner, who was the world's leading antitrust lawyer at the time, was the head of the Antitrust Division. Edwin Zimmerman, who was an extraordinary talent, was the Deputy Assistant Attorney General for Antitrust. Stephen Breyer was Turner's Special Legal Assistant. And Richard Posner was arguing antitrust cases before the Supreme Court as a lawyer on the staff of the Solicitor General.
10. Also note that the “habit of breaking down problems into smaller parts” promotes discovery (Hoddeson and Daitch, 2002, p. 316). Also see Simon (1962) on the

- importance of near-decomposability.
11. Getting it right in a logical sense does not imply that the theory is correct. As Friedman observes, "Science in general advances primarily by unsuccessful experiments that clear the ground" (in Snowdon and Vane, 1992, p. 296). That is useful to bear in mind. Most would-be theories fail – which explains why Simon favors pluralism: perhaps someone will get it right (Simon, 1999, p. 22).
 12. Solow maintains that plausible simple models of complex phenomena are expected to “make sense for 'reasonable' or 'plausible' values of the important parameters” (Solow 2001, p. 112). The plausibility precept challenges users of fanciful constructions to explain “What is going on here?”
 13. Personal communication from Milton Friedman to the author, Feb. 6, 2006.
 14. I discussed similar difficulties with Edward Prescott (who was also in the PhD program at Carnegie in the 1960s). On both occasions Ed advanced the view that the economics profession was in need of a “new mathematics’ to better address governance issues. Be that as it may, it is noteworthy that full formalism efforts that originated with Sanford Grossman and Oliver Hart (1986) and are works-in-progress to this day (see Section 4 of Tadelis and Williamson, 2012).
 15. For a more extensive version of the Peguy quote, see Williamson (1996, p. 13).
 16. This article was instrumental in promoting empirical analysis across a wide range of fields in economics, business, and the contiguous social sciences. (It was reprinted 16 times.) Although credible commitments are sometimes thought of as a user-friendly way to contract, credibility is actually hardheaded in that it is used

- in cost effective degree to support those transactions where asset specificity and contractual hazards are at issue. Such supports are without economizing purpose for transactions where generic technologies are employed.
17. TCE addresses the issue of scaling up in Williamson (1985, pp. 96-98). The text here is based on Tadelis and Williamson (2012).
 18. Michael Jensen and William Meckling expressly recognized the importance of scaling up from a single owner-manager to a multitude of owners in a modern corporation and stated that this was an issue that they planned to deal with it in a later paper (1976, p. 356). That paper never materialized, presumably because of the difficulties. (Their 1976 paper has nonetheless been influential.)
 19. The Institutional Environment branch of Institutional Economics is crucial. The work of Douglass North and his colleagues is especially important in this connection.
 20. *International Herald Tribune* (1990; italics added).
 21. As reported in Williamson (1976, p. 91, n. 11).
 22. As reported in Williamson (1976, pp. 92-101).
 23. See Williamson (1985, p. 199) for more such “reinterpretable” statements.
 24. For citation counts of each, see Williamson (2005b).
 25. George John and Torger Reeve especially emphasize the 1979 article which “spurred empirical work” in marketing by dimensionalizing the attributes of transactions and governance structures and advancing “refutable conjectures” (2010, p. 249).

26. Such tautological reasoning gave transaction cost a bad name in its early years, which critique also applies to many users of the term “power”—which is a word of many meanings, as witness that it takes up over an entire column in Webster's Third New International Dictionary. If, therefore, power is used by social scientists, the delimited purpose intended ought to be disclosed.

Although some social scientists contend that while power is “tricky to define ... it not that difficult to recognize“ (Pfeffer, 1981, p. 3), but this repeats the confusion that is associated with the word trust. As James March puts it, power is a “disappointing concept. It tends to become a tautological label for the unexplained variance“(1988, p.4)

Transaction cost economics concurs (1985, pp. 238-239).

27. See p. 49 of the text.

28. The diagrams and comparisons track Tadelis and Williamson (2012).

29. This oversimplifies but is broadly congruent with the spirit of simple market exchange. Courts, for example, sometimes award other relief, such as specific performance. But “classical contract law“ (Macneil, 1978, pp. 862-864), which applies to the polar transaction of $k = 0$, is basically a money damages transaction.

30. My suggestion is that such a research program would benefit by working through the voluminous bureaucracy literature in the classroom. This is most certainly a case where “teaching is learning.”

31. Additionally, the distinction that TCE makes between perfunctory and consummate cooperation in the context of ongoing contractual relationships is important and stands in need of further development.

32. For a recent discussion of and contribution to full formalism, see Tadelis and Williamson (2012).