The Transaction Costs Paradigm:

1998 Presidential Address Western Economic Association

Steven N. S. Cheung

Though the idea can be traced to David Hume and Adam Smith, the path-breaking first economic analysis of transaction costs did not occur until 1937, when Ronald Coase published "The Nature of the Firm."1 Stigler and Boulding perspicaciously reprinted Coase's paper in the famous AEA Readings in Price Theory {1952}, but for a quarter of a century it had no real impact on the profession. In the 1940s two significant works were published which touched on transaction costs, but these sparks did not ignite a fire.2

The transaction costs paradigm began to blossom in the early 1960s. Again it was Coase who supplied the primum mobile-his great paper on social cost was published in early 1961, though Aaron Director dated that issue of Journal of Law and Economics as 1960. Ronald had company this time: George Stigler's article on information cost came out later the same year, and Kenneth Arrow's work on the appropriability of returns to inventions appeared in 1962.3

The combined wisdom of three brilliant men would perhaps not have been enough, except that by the early 1960s science was on their side. Jargonistic articles on economic development flooded the journals throughout the 1950s, but by 1960 economists were beginning to question the explanatory power of this kind of economics. After the clarion call of Friedman's work on the consumption function (a few years earlier, in 1957), the great methodological debate began.4 The demand for testable implications was becoming more vociferous by the day, and the tendency for economists to be policy oriented was losing hold as the Keynesian revolution waned.5 The tide of economic explanation was beginning to swell, and the transaction costs paradigm was the beautiful goddess who emerged from the foam.

I began graduate work at UCLA in 1961, and after my MA a year later I turned to audit the price theory lectures of Jack Hirshleifer and Armen Alchian. The straight As student

who knew by heart everything about market imperfections was told that this kind of knowledge did not count!

By 1963, graduate students at UCLA's Department of Economics were all talking property rights and transaction costs. Now I knew Coase and Stigler by heart, and could even handle the questions on pricing behavior and mergers at the end of Friedman's Price Theory {1962}. Moreover, I was the luckiest of the lot because in late 1963 Armen gave me a copy of a manuscript by Harold Demsetz.6 One of the great expositors of our time, Harold wrote about property rights and transaction costs with such clarity and force that I decided then and there to work on these topics for my doctorate. Since that time, my research has all been on the same track.

In this address, I shall confine discussion of the transaction costs paradigm to my own involvement. This limitation is imposed by necessity. I like to sail the strange seas of thought alone, and have seldom read other people's publications with the seriousness they deserve after I left Chicago in 1969. From 1976 to 1982, I spent my time on contracts in the petroleum industry, and from 1982 to now, my research has concentrated on economic reforms in China. Although these works are all transaction-costs oriented, the former is not in print because it is proprietary, while the latter has mostly been published in Chinese. After more than half a lifetime on the seashore, however, I feel it is time to share my small collection of pebbles with colleagues in the WEA. Still, you would not be wrong if you feel that I am addressing the subject as if I were a stranger returning from another planet.

I. THE NATURE OF TRANSACTION COSTS

"Transaction costs" must be defined to be all the costs which do not exist in a Robinson Crusoe economy. This broad definition is necessary, because it is often impossible to separate one type of transaction cost from another: the toll collector of a bridge serves not only transacting customers but also police against trespassers. What is important is that inseparability does not hinder the derivation of testable hypotheses. Indeed, it is isomorphic to the case of joint production with variable proportions where average costs are not separable but marginal costs are, and so long as marginal costs are separable testable hypotheses may follow.7

I have suggested, with the full approval of Coase, that transaction cost should actually be called "institution cost." An economy of more than one individual would necessarily contain institutions,8 but the costs that arise as a result may entail no transactions at all. For example, during the cultural revolution in China there were hardly any transactions in the market, but the political costs of memorizing Mao's slogans, establishing connections and so on were enormous. These costs certainly cannot exist in a Robinson Crusoe economy. They arise only where there are institutions, or in a "society" in the plain sense

of the term. But changing household terminology is nearly impossible, so "transaction costs" stays even though it is not strictly correct and may even be misleading.

"Transaction costs" thus broadly defined are huge indeed. Variations cover the incomes of lawyers, financial institutions, policemen, middlemen, entrepreneurs, managers, clerks, civil servants, ... just about all the conceivable costs in society except those associated with the physical processes of production and transportation. In today's Hong Kong, for example, when nearly all factories have moved north to the mainland, at least 80% of GDP derives from transaction costs, mainly as a result of servicing economic activity in China. In agricultural and non-politicized countries, transaction costs as a proportion of income would of course be less. But in the modern world, it would be difficult to find a rich country where transaction costs sum to less than half of national income.

A question that must then be raised is why, given the magnitude of transaction costs, the subject could have been ignored by economists for so long? I believe there are two reasons.

The first is that before the 1960s, with the single exception of Ronald Coase, economists tended to think of transaction cost as similar to transportation cost, or similar to a tariff or a commission. Students of my generation are all too familiar with the empty boxes we played with. Since Marshall's time economists did not, and many still do not, like to tackle a problem unless it fitted nicely into geometry or algebra or calculus. As traditionally understood, transportation cost, tariffs or commissions would not produce observations other than those pertaining to resource allocation and income distribution. For analytical convenience, therefore, deleting such price components would involve nothing more than a simplifying assumption. Thus it transpired that the followers of Walras adopted the convention, under which the auctioneer supplies services free of charge.

The truth of the matter, of course, is that transaction cost is not the same thing as transportation cost. Changes in transaction costs, in one dimension or another, would generally lead to changes in the contractual or organizational structure. This is so because it may be possible to reduce transaction costs by rearranging institutions: the society we live in and the way we conduct economic activities depend upon the magnitude and type of cost which govern institutions in its numerous forms. You may recall that the Walrasian equations proceed in terms of n commodities and n-1 relative prices. The glaring hiatus in this approach is that the number n cannot be determined without introducing transaction costs, which however Walras assumed away.

A second reason for economists' long neglect of transaction costs is that the concomitant constraints cannot be properly specified without knowing a lot about what is going on in the real world. This, unfortunately, is routinely difficult and exhausting. Transactioncost

economics is real-world economics, and the real world is too often a place where academic economists fear to tread. Who would want to commit two years to studying an antitrust case, with an uncertain prospect of eventual publication? Pauca sed matura is a motto only a Gauss can afford.

It was most unusual that it took only three months for me to complete an investigation of the constraints underlying beehive rentals, and write a paper everybody was keen to print.9 I also had luck with landlords and tenants (four years)" and petroleum contracts (six years). But in patent and trade secret licenses, I ended up almost empty-handed after five years of trying. Transaction costs is not a subject a young economist seeking tenure should go into!

It has been argued that it is fruitless to study transaction costs, because it is frequently impossible to measure them. This view is wrong. Fundamentally, measurement involves an assignment of numbers for the purposes of ranking, and precision in measurement can only be judged by the extent of agreement among different observers. To say that cost is measurable, or measurable precisely, does not necessarily mean it is measurable in dollars and cents. If we are able to say, ceteris paribus, that a particular type of transaction cost is higher in Situation A t han in Situation B, and that different individuals consistently specify the same ranking whenever the two situations are observed, it would follow that transaction costs are measurable, at least at the margin. Testable propositions may then be obtained, and that is the important thing.

II. SPECIALIZATION WITH TRANSACTION COST CONSTRAINTS

Ricardo and Mill were only half right in their conclusions about specialization and comparative advantage. Comparative advantage, no doubt, promotes specialization, but it is not a necessary condition for specialization to occur. Even if everybody were born with identical genes so that we all possessed equal natural advantages, I doubt that there would be very much less specialization in the world. Rather, specialization sharply reduces the cost of learning. In addition to comparative advantage, the gain from specialization would therefore far exceed what Adam Smith envisioned with his story of the pin factory.

An example may illustrate the point. A ball-point pen has a retail price of 25 cents, and of this sum the manufacturer receives no more than a nickel. The pen consists of metal, plastic, petrochemicals, in addition to design and the great invention of the rolling ball. If none of these materials and technologies were known and one were required to manufacture a ball-point pen from scratch, the cost would be at least one billion times higher.

Gains from specialization in Adam Smith's time were nowhere near as large as they are now. Specialization not only goes hand-inglove with physical production, but more

importantly, it does so in the production of ideas. Ideas and innovations are indestructible and accumulate over time; they are also amenable to concurrent usage. Specialization in production reinforced with innovations fully explains why Malthus was miles off in his dismal prognosis of human livelihood.

Adam Smith was also half right in saying that specialization is limited by the extent of the market. As I will elaborate below, though specialization requires the support of exchange, market transactions merely comprise one form of exchange. Because the gains from specialization are enormous, there is room to accommodate very large transaction costs from exchange in all of its variations, and still have resources a-plenty left over for all to enjoy. In this way, transaction costs as a proportion of gains from specialization provides a critical measure which, in my view, very largely explains observed differences in the wealth of nations.

Economists have long been baffled by these differences. In particular, they have been intrigued by the fact that some societies have managed to do extremely well despite poor endowments in natural resources. The argument that different economic systems--ifferent property rights systems-matter is of course correct, but it is incomplete. What is essential is the recognition that under different systems transaction costs differ as a proportion of gains from specialization, and if this ratio is reduced just a little, a significant increase in wealth would follow.

My favorite example is China under communism, especially during the Great Leap Forward and the Cultural Revolution. There was hardly any market, but exchange did occur through a central distribution system. With exchange there would be specialization, but to realize the gains one had to contend with huge transaction/institution costs, leaving very little afterwards. In other words, under the communist regime transaction costs as a proportion of gains from specialization was very large. This, I believe, is why the Chinese were so poor in their communist days.

III. REINTERPRETATION OF THE PARETO CONDITION AND THE DISSIPATION OF RENT

It was Coase who suggested that in the presence of transaction costs Pareto optimality requires reinterpretation;ii Demsetz reinterpreted it brilliantly;12 I went so far as to say that in the real world the Pareto condition is always satisfied.l3 Under the postulate of constrained maximization, in a one-man system economic inefficiency is, by definition, impossible. Robinson Crusoe may starve to death because he is constrained by incompetence, but there is nothing "inefficient" about it. The riddle then is why economists have been so prone to suggest the frequent occurrence of inefficiency in a society, in which Pareto's condition is claimed to be violated? My answer is that a violation of this nature can occur only if certain constraints are neglected or ignored.

A self-service buffet dinner provides a good illustration. An individual paying a lump-sum and free to eat as much as he pleased will consume to the point where the marginal benefit from the last mouthful reaches zero. The marginal cost of producing that last mouthful, however, is positive. Conclusion: the Pareto condition is violated. However, the picture changes if we ask why the buffet dinner is served as such. The answer, of course, is that it saves the costs of waiting upon customers and of metering the food each consumes. These transaction-cost savings must be larger than the "waste" generated by on the limit consumption, and once we take them into account the Pareto condition would be satisfied.

Here there are two observations which require explanation. One is the eating behavior of buffet customers; the other is the choice of the buffet dining contract. To explain the former, there is no need to introduce the costs of serving and metering, and this omission leads to a conclusion of economic waste. To explain the latter serving and metering costs must be brought in, after which the waste disappears. Some economists tend to think that if the buffet customers behaved themselves and ate Pareto-optimally, a great deal of food will be saved and the price of dinners would accordingly be reduced, to the benefit of all. How nice would it be if the restaurant owner merely posted the marginal costs on the wall, and all customers would follow by eating equi-marginally? Yes, Sir Thomas More, the world would be a better place if we do not shirk, cheat, lie, or steal.

There is no doubt transaction costs would be far lower if we were all born with decent genes, as if we were congenitally disposed to obey the Ten Commandments, and we would all be richer as a result. But "indecent" behavior is the result of constrained maximization, and this postulate must be asserted as universal or there would be no economic science. Selfishness helps society; selfishness also harms society. We seek to reduce transaction costs, but we also behave to increase them. Under different institutional arrangements the distribution of these gains and costs will differ. Indeed, the Ten Commandments can be interpreted as an institution designed to reduce transaction costs.

In the buffet dinner example, it is clear that the Pareto condition appears to be violated only because certain transaction cost constraints are omitted, and because it is not necessary to introduce such constraints to explain a certain kind of behavior. Specification of the constraints sufficient for interpreting a given observation may not suffice to produce an "efficient" outcome, but this must not be construed to mean that the Pareto condition would remain violated when all transaction/institution costs are taken into account.

The buffet dinner argument may be generalized to all "inefficient" activities in society. Take price or rent controls, which yield the standard examples of economic wastage. Here again, transaction/institution costs which underlie the political and legislative

processes of control have been ignored. If these costs are brought fully into consideration, the Pareto condition would be satisfied.

My reinterpretation of Pareto optimality renders the condition worthless in welfare economics, but significantly enhances its role in positive analysis. In specifying constraints to derive testable propositions, whenever the Pareto condition fails to hold we would immediately know that some constraints are missing: it would then be up to us to decide whether the omitted constraints are relevant to the observations we are seeking to explain.

Equilibrium, defined as a state in which there is sufficient specification of constraints to yield testable implications, must display transaction costs (if relevant) as a constrained minimum. Failure to do so would render the analysis empirically empty. A notable example is the literature on the dissipation of rent, in which rent without exclusive claimants is said to be dissipated to produce an equilibrium. Often what replaces the rent dissipated are transaction costs in some form or another. If these are not treated in terms of constrained minimization, we cannot find a way to predict the particular behavior through which rent would dissipate.

Treating transaction costs as a constrained minimum is a requisite for treating dissipation of rent as a constrained minimum. This is absolutely essential, if the postulate of constrained maximization is to be consistently applied. As I have noted in a study of price control, one cannot predict whether it would lead to queueing, politicking, favor-currying, or violence, unless the dissipation of rent is treated in terms of constrained minimization. 14

My prolonged research on China's economic reforms has added another insight to this view. In the real world, there is no such thing as a valuable resource open to unrestrained common exploitation. When private property rights are absent, some other rights must emerge to fill the void. These "other rights" are in a substantive sense exclusive, whether they are assigned by hierarchical ranking, political connection, or sex appeal. Which way depends on transaction/institution costs, in a manner that keeps the dissipation of rent to a constrained minimum.

Under its earlier communist regime, China chose a complicated but ingenious hierarchical ranking system to assi gn rights over resources. Hunger was widespread, and yet the population exploded. This paradox may be explained by noting that hierarchical ranking is an effective way to reduce the dissipation of rent.

IV. A FALLACY IN THE COASE THEOREM AND THE THEORY OF THE STATE

What would things be like if transaction/institution costs were zero? There are a number of possibilities, but one thing we can be certain of is that in such a world property rights or institutional arrangements would not matter. Analyzing institutional change in China, I wrote in 1982:15

If all transaction costs, broadly defined, were truly zero, it would have to be accepted that consumer preferences would be revealed without cost. Auctioneers and monitors would provide free all the services of gathering and collating information; workers and other factors of production would be directed freely to produce in perfect accord with consumer preference; and each consumer would receive goods and services in conformity with his preferences. The total income received by each worker (consumer), as determined costlessly by an arbitrator, would equal his marginal productivity plus a share of the rents of all resources according to any of a number of criteria costlessly agreed upon. In other words, production and consumption activities can in principle be carried out without a market, to produce the same result as though costless markets were in operation.

An important implication follows: the market itself is an institution, which would not have emerged if transaction/institution costs were zero. Like any other institution, the market was created to reduce transaction costs, subject to other constraints.

Now the Coase Theorem, in its earlier FCC version, states that private property rights is a necessary condition for the theorem of exchange to be operative. In Coase's own words: "The delimitation of rights is an essential prelude to market transactions."16 In essence this is an important and fruitful way to restate and apply the theorem of exchange, in which I can find no flaw.

On the other hand, I am disturbed by the Coase Theorem in its later social cost incarnation. This popular version, sometimes referred to as the invariance theorem, states that if private property rights exist and if transaction costs are zero then resource allocation will be "invariantly optimal," regardless of how the rights are assigned. The Coase Theorem so stated must be fallacious, because logically private property rights cannot coexist with zero transaction costs.17

Private property rights is itself an institution and, like any other institution, it arises because transaction costs are not zero. As with any institutional arrangement, choosing a system of private property rights depends on how far transaction costs are reduced subject to other constraints. This brings to mind an observation made by Friedman some forty years ago, when he said it is foolish to try to empirically estimate whether a firm's cost of production is the lowest, for it must always be so by definition.18 Similarly, if an institutional arrangement is observed to lead to massive starvation, by definition the tragedy is the result of constrained maximization. The problem of economic science is to begin with such a tautology, and then to derive falsifiable propositions to test the

explanation of what appears to be a stupid choice. In the process, we should remember that tautologies, assertions and identities always appear at the starting point of scientific analysis.

When in 1966 I conceived the notion that contractual arrangements are determined by choice subject to transaction cost constraints,19 I thought I was on to something novel. A few months later I realized I was mining the same vein as Aaron Director's oral tradition on tie-in sales, and Ronald Coase's work on the firm. I did not know it when I worked on sharecropping for my doctoral thesis, but in retrospect there is no way I could have completed the thesis without Aaron's and Ronald's influence.

There is no reason why we cannot or should not extend the analysis of pricing and contractual arrangements to problems of larger scale, in particular by treating the constitution of a country as a contract and the state as a giant firm.

In 1981, employing a transaction costs approach, I predicted that China will go capitalist.20 Many friends demurred-some (particularly Ted Schultz) even argued that economic analysis cannot be used to make this type of prediction-but I let the manuscript go into print because I was convinced the analysis was correct. In this I owe a great debt to my old friend and colleague Yoram Barzel. Yoram too was skeptical, but he was also firm that my reasoning was flawless and urged that I publish. Judging from what has happened in China, it is clear that we have here at least one example, which shows how changes in transaction cost constraints may be applied to explain the evolution of the state.

V. CONCLUDING REMARKS

Economic analysis seems to have become a great deal more complicated since my student days. I find this development disturbing. The world being such a complicated place, we would stand little chance of interpreting its institutional arrangements if complicated tools are used. The transaction costs paradigm in which I was brought up-and here I am sure Coase fully shares my view-has the merit that it entails only the simplest of economic tools. In fact, this paradigm contains no new theory whatsoever to speak of.

Only three fundamental propositions are present in the paradigm. First is the postulate of constrained maximization. Second is the downward sloping demand curve, which (because there is no need to separate consumption and investment activities) also covers diminishing marginal productivity. Third is the notion that cost is the highest-valued option foregone.

In this tradition, the transaction costs paradigm concentrates on changes in constraints. Nothing is new in the theory, and the analytical tools are kept to the most elementary.

The paradigm, however, is simple but difficult. The difficulty lies in the thoroug empirical investigation required, from which we can (hopefully) garner insights on the nature and classification of transaction cost constraints in the real world.

On a fundamental level, there are only three avenues through which we can obtain economic interpretation of observations. First is manipulating the utility function; second is manipulating the production function; third is manipulating the constraints. (Combinations are of course possible.) New theory may emerge along the first and second avenue, but never with the third. My stance, and no doubt it is a minority view, is that institutional economics should confine itself to the third avenue.

Interpretations of institutional arrangements have been attempted through all three avenues, together with a variety of combinations. I harbor mixed feelings about a major work, to which I contributed orally. This is the now famous Alchian-Demsetz paper on economic organization.21 Two of the leading price theorists of our generation, Armen and Harold combined manipulations of transaction cost constraints as well as the production function. Their paper has become the headwater of the game theoretic approach to contracts. This paradigm is too complicated for my liking. In my view we still have to wait, to see what predictive powers the fashionable high theory may or may not display.

1. Coase {1937}.

2. See Hayek {1945} and Coase {1946}.

3. See Stigler {1961} and Arrow {1962}.

4. It was Alchian's classic article {1950} which started the whole thing, and after Friedman's equally classic {1953} defence of the use of behavioral assumptions in positive economics, pundits like Gordon {1955}, Becker {1962}, Nagel {1963} and Samuelson {1963} entered the fray.

5. Robbins {1961} observed that classical economists were all policy-oriented. This tradition was continued by Pigou and Keynes, but not by Fisher and Knight.

6. I believe this manuscript was later published as two papers, at the cost of some loss in forcefulness. See Demsetz {1964} and {1967}.

7. Hirshleifer suggests that it is essential to distinguish between two types of transaction costs: those associated with "exchanges", and those associated with "commands from higher authority". This observation would seem to be in the same vein as Coase's {1937} treatment of the firm and the market. However, the inseparability problem does not

disappear with this distinction. Indeed, I have argued that often there is no clear dividing line between the firm and the market. See Cheung {1983}.

8. The word "insitution" may seem vague, but I use it to simply mean any arrangement used to conduct economic activities which involve two or more individuals.

9. See Cheung {1973}. When they saw the circulated manuscript, two editors (including the AER's) wrote to ask to publish it. But the paper was committed to Coase even before it was written down.

10. See Cheung {1974}, {1975} and {1979}.

11. Coase {1946} and {1960}. 12. Demsetz {1969}.

13. Cheung {1974} and {1982}.

14. See Cheung {1974}. 15. Cheung {1982}.

16. Coase {1959, p. 27}.

17. This argument promises to engender controversy in the future literature, because even at this stage Hirshleifer and Barzel have difficulty with it. I, of course, believe that my views will stand the test of time and debate.

18. See Friedman {1962, p. 139 and if}. 19. See Cheung { 1969}. 20. Published as Cheung {1982}.

21. Alchian and Demsetz {1972}.

REFERENCES

Alchian, A. A. "Uncertainty, Evolution, and Economic Theory." Journal of Political Economy, June 1950, 211-21.

Alchian, A. A. and H. Demsetz. "Production, Information Costs and Economic Organization." American Economic Review, December 1972, 777- 95. Arrow, K. J. "Economic Welfare and the Allocation of Resources for Invention," in National Bureau of Economic Resea rch. The Rate and Direction of Inventive Activity. Princeton: Princeton University Press 1962, 609-25.

Becker, G. S. "Irrational Behavior and Economic Theory." Journal of Political Economy, February 1962, 1-13. Cheung, S.N.S. "Transaction Costs, Risk Aversion, and the Choice of Contractual Arrangements." Journal of Law and Economics, April 1969, 23-42. . "The

Fable of the Bees: An Economic Investigation." Journal of Law and Economics, April 1973, 11-33.

. "A Theory of Price Control." Journal of Law and Economics, April 1974, 5371.

* "Roofs or Stars: the Stated Intents and Actual Effects of a Rents Ordinance." Economic Inquiry, March 1975,1-21.

. "Rent Control and Housing Reconstruction: the Postwar Experience of Prewar Premises in Hong Kong." Journal of Law and Economics, April 1979, 27-53.

. Will China Go "Capitalist"? London: Institute of Economic Affairs, 1982.

. "The Contractual Nature of the Firm." Journal of Law and Economics, April 1983.

Coase, R. H. "The Nature of the Firm." Economica (NS), November 1937, 386-405.

"The Marginal Cost Controversy." Economica (NS), August 1946,16982.

"The Federal Communications Commission." Journal of Law and Economics, October 1959, 1-40. . "The Problem of Social Cost." Journal of Law and Economics, October 1960,1-4.

Demsetz, H. "The Exchange and Enforcement of Property Rights." Journal of Law and Economics, October 1964, ll-26.

. "Towards a Theory of Property Rights." American Economic Review, May 1967, 347-59.

"Information and Efficiency: Another Viewpoint." Journal of Law and Economics, April 1969,1-22. Friedman, M. "The Methodology of Positive Economics." in idem, Essays in Positive Economics. Chicago: University of Chicago Press, 1953. . A Theory of the Consumption Function. Princeton: Princeton University Press, 1957. . Price Theory: a Provisional Text. Chicago: Aldine Publishing Company, 1962.

Gordon, D. F. "Operational Propositions in Economic Theory." Journal of Political Economy, April 1955, 150 61.

Hayek, F. A. "The Use of Knowledge in Society." American Economic Review, September 1945, 519-30. Nagel, E. "Assumptions in Economic Theory." American

Economic Review, May 1963, 211-19. Robbins, L. C. The Theory of Economic Policy in English Classical Political Economy. London: Macmillan and Co., 1961.

Samuelson, P. A. "Problems of Methodology-Discussion." American Economic Review, May 1963, 23136.

Stigler, G. J. "The Economics of Information." Journal of Political Economy, June 1961, 213-25.

* I am indebted to M.T. Cheung for assistance with the final four of seven drafts of this paper, and for this Michael was awarded the original handwritten draft. I have, after all, never touched a word processor. Among old friends and peers Milton Friedman, Jack Hirshleifer and Yoram Barrel all rushed to my aid-Milton by fax, Jack by e-mail, and Yoram by hand.
