

TRANSACTION COSTS THROUGH TIME

An economic definition of transaction costs are the costs of measuring what is being exchanged and enforcing agreements. In the larger context of societal evolution they are all the costs involved in human interaction over time. It is this larger context that I wish to explore in this essay. The concept is a close kin to the notion of social capital advanced by James Coleman (1990) and applied imaginatively to studying the differential patterns of Italian regional development by Robert Putnam in Making Democracy Work (1993). This essay, therefore, is a study in economic history which focuses on the costs of human coordination and cooperation through time which I regard as the key dilemma of societies past, present and future.

The evolution of societies is a function of the quantity and quality of human beings, the human command over nature, and the structure humans impose on their interaction. An understanding of the interaction between demographic, technological, and institutional factors would provide fundamental insights into societal evolution. This essay is a very preliminary step at exploring the transaction costs involved in the interaction of two of these three basic determinants societal evolution, institutions and technology.

Three landmarks in the historical reduction of transaction costs were the institutions that made possible impersonal exchange, the assumption by the state of the protection and enforcement of property rights and realization of the gains from the modern revolution in science. In what follows I shall begin by posing the dilemma of human coordination and cooperation in terms of standard transaction cost analysis (I), explore the way the dimension of time affects the costs of transacting (II). examine some of the major changes in economic institutions that laid the foundations of impersonal exchange(III). Section IV explores the conditions that underlay the assumption by the state of the protection and enforcement of property rights. I then analyze the complex interplay between technological and institutional change in the past century and a half that arose from the revolutionary changes in scientific knowledge(V).

I

Ronald Coase forced economists to think about the costs involved in human interaction. He was concerned to explain the reason for the existence of firms (1937) or the conditions under which the allocative implications of microeconomic theory held (1960). But the study of transaction costs in addition to giving us insights into static economic analysis also holds the key to unlocking the doors to an improved understanding of economic and societal performance through time.

In the ten millenia since the first economic revolution humans have haltingly evolved institutions to structure human interaction that have permitted and encouraged increasing productivity and economic growth. Broadly speaking the economic institutions have been those that have permitted the growth of markets or improvements in, or the introduction of, new technology. The political institutions have been those that improved the security of property rights and the enforcement of contracts. In fact the two

institutional sources have been inextricably intertwined. The historical decline in transaction costs has reflected both voluntaristic and coercive solutions to problems of exchange. Since I have outlined this evolution elsewhere I shall not repeat it here.¹ What I wish to explore here is innovations that permitted such significant reductions of transaction costs that they made possible production and exchange that had not existed before. I wish to do this in order to explore the specific conditions in time that led to these breakthroughs. But first I must look at the way learning through time determined these changes.

II

In economic analysis change is primarily a consequence of changes in relative prices. And indeed that has been a powerful tool to explore historical change.² Here I seek to understand the underlying sources of changes in relative prices. Many of them arise from exogenous shocks or the gradual accumulation or diminution of one supply factor more rapidly than another. It has been technological change, however, that has been the primary explanatory variable of economists and economic historians. But clearly that explanation is incomplete. We are not only left with the question of what has determined the varying rate of technological change through history but also with the question of what has determined the differing ability of economies to take advantage of the technology that exists. Surely if technology was the whole answer all the world should be rich since most of the technology is openly available. The inability of economies to make use of the existing technology efficiently suggests that it is the (dis)incentives embodied in economic and political organization that is at issue. That in turn leads us to the belief systems that humans hold which shape the institutions that they create.

Neo-classical theory assumes that people know what is in their self interest and act accordingly. This rationality tenet of faith is simply incorrect as a guide to the perceptions of humans throughout history that have shaped the institutions that they have created to structure human interaction. Instead the learning process that has shaped the belief systems of humans throughout history has produced widely diverse perceptions about how the world is and should be ordered.³ Most have led to poor economic performance or stagnation; a few have led to economic growth.

Human learning is a product of the accumulated experience of past generations carried over intergenerationally as culture and the "local" experiences of the members of a society.⁴ As societies evolved from tribal beginnings they developed different languages and, with different experiences, different mental models to explain the world around them. With growing specialization and division of labor they evolved polities and economies but the diversity of experiences and learning produced increasingly different societies with

¹. See the essay on "Institutions" in the Journal of Economic Perspectives, Winter, 1991

². See North and Thomas, 1973 and North, 1981.

³. For a discussion of cognitive science and the foundations of human learning see Denzau and North, 1994 and North 1994

⁴. See Denzau and North, 1994 for an elaboration of this argument

different degrees of success in solving the fundamental problems of scarcity. The reason is that as the complexity of the environment increased with increasing interdependence institutions were required to permit anonymous, impersonal exchange and to structure polities to provide protection and enforcement of property rights but the likelihood of creating the necessary institutions to capture the gains from trade of more complex contracting varied. In fact most societies throughout history got stuck in an institutional matrix that did not solve the institutional requirements involved in the three major transaction cost reducing innovations described at the beginning of this essay.

III

Between the 11th and 14th centuries in Europe a commercial revolution led to the revival of trade and an era of substantial economic growth. The key to this expansion was the development of a set of institutions that permitted anonymous exchange to take place across space and time. Inter-community credit markets, insurance markets, contracts for future delivery, and the bill of exchange all were institutional features of this commercial revolution.

The evolution of the Law Merchant was characteristic of the institutions that undergirded this expansion. Merchants gradually evolved codes of conduct in different parts of western Europe to define exchange relationships among themselves. In the absence of a coercive enforcement mechanism by states reputation and ostracism served as the basis of enforcement. But at the size of the markets grew a reputation mechanism, alone, was an insufficient guarantee of performance and a legal code administered by private judges drawn from commercial ranks enhanced the effectiveness of the reputation mechanism by providing incentives for information dissemination, imposition of sanctions on violators, and payment of judgements levied against an individual merchant.⁵ A critical feature of the Law Merchant is that it evolved into and was integrated with formal legal codes downstream thus providing a path dependent evolution from informal codes of conduct to formal coercive enforcement by the state.

This last point deserves special emphasis. Developing institutions to structure impersonal exchange occurred in many economies during the era of the commercial revolution. But many such innovations did not, in contra distinction to the Law Merchant, lead downstream to further institutional development. Some dead ended with no further development. Avner Greif in a forthcoming study compares Genoese traders with traders who had adopted the cultural and social attributes of Islamic society in the Mediterranean trade of the 11th and 12th centuries. He detects systematic differences in their organizational structure traceable to contrasting individualistic versus collectivist behavioral beliefs. Traders from the Islamic world developed ingroup social communications networks to enforce collective action which, whole effective in relatively small, homogeneous ethnic groups, do not lend themselves to the impersonal exchange that arises from the growing size of markets and diverse ethnic traders. In contrast the Genoese developed bilateral enforcement mechanisms which entailed the creation of formal

⁵. For an analysis of the institution and a game theoretic model of the way it worked see Milgrom, North, and Weingast, 1990

legal and political organizations for monitoring and enforcing agreements--an institutional/organizational path that permitted and led to more complex trade and exchange. Greif suggests the generality of these different belief structures for the Latin and Muslim worlds and then makes the connection between such belief systems and the subsequent institutional development of the western world that led to modern economic growth.⁶

IV

It was in western Europe in general and the Netherlands and England specifically where polities emerged that took over the protection and enforcement of property rights. It was the lack of large scale political and economic order that created the essential environment hospitable to political/economic development. In that competitive decentralized environment lots of alternatives were pursued as each society confronted its own unique external environment. Some worked as in the case of the Netherlands and England; some failed as in the case of Spain and Portugal. But the key to the story is the variety of options pursued and the likelihood that some would turn out to produce political/economic development. Even the relative failures in Western Europe played an essential role in European development and were more successful than other parts of the world because of competitive pressures.

The last point deserves special emphasis. It was the dynamic consequences of the competition amongst fragmented political bodies that resulted in an especially creative environment. Europe was politically fragmented; but it was integrated in having both a common belief structure derived from Christendom, and information and transportation connections that resulted in scientific, technological, and artistic developments in one part spreading rapidly throughout Europe. To treat the Netherlands and England as success stories in isolation from the stimulus received from the rest of Europe (and to a lesser degree Islam and China) is to miss a vital part of the explanation. Italian city states, Portugal, and Germanic states all fell behind the Netherlands and England; but banking, artistic development, improvements in navigation, and printing were just a few of the obvious contributions that the former states made to European advancement.

The Netherlands and England pursued different paths to political/economic success but in each case the external environment was conducive to the evolution of a belief structure that induced political and economic institutions that lowered transaction costs.

In both polities competition among the evolving nation states was a deep underlying source of change and equally a constraint on the options available to rulers within states. It was competition that forced the crown to trade rights and privileges for revenue including most fundamentally the granting to representative bodies--variously Parliament, States General, Cortes--control over tax rates and/or certain privileges in return for revenue. But it was the evolving bargaining strength of rulers vis-a-vis constituents that was the decisive feature of their subsequent development. Three considerations were at stake: 1) the size of the potential gains the constituents could realize by the state taking over protection of property; 2) the closeness of substitutes for the existing ruler--that is the ability of rivals (both within and outside the political unit) to the existing ruler to take over and provide the same, or more, services; 3) the structure of

⁶. Greif, Avner, "Cultural Beliefs and the Organization of Society"

the economy which determined the benefits and costs to the ruler of various sources of revenue.

Let me briefly describe the background conditions of the two polities--the Netherlands and England--that led up to the contrasting external environments that shaped the belief systems.

To understand the success of the Netherlands one must cast a backward glance at the evolution of the prosperous towns of the Low Countries such as Bruges, Ghent, and Liege; their internal conflicts; and their relationship to Burgundian and Habsburg rule. The prosperity of the towns, whether based on the wool cloth trade or metals trade, early on made for an urban centered, market oriented area unique at a time of overwhelmingly rural societies. Their internal conflicts reflected ongoing tensions between patrician and crafts and persistent conflicts over ongoing efforts to create local monopolies which, when successful, led to a drying up of the very sources of productivity which had been the mainspring of their growth. Burgundian (and later Habsburg) rule discouraged restrictive practices such as those that developed in the cloth towns of Bruges and Ghent and encouraged the growth of new centers of industry that sprang up in response to the favorable incentives embodied in the rules and property rights. In 1463 Philip the Good created a representative body, the States General, which enacted laws and had the authority to vote taxes for the ruler. The Burgundians and Habsburgs were rewarded by a level of prosperity that generated tax revenues that made the low countries the jewel in the Habsburg Empire.

England evolved along a route different from that of continental polities. Being an island made it less vulnerable to conquest and eliminated the need for a standing army (and undoubtedly contributed to the different initial belief structure that Macfarlane describes). The Norman conquest, the exception to British invulnerability to external conquest, produced a more centralized feudal structure than on the continent. The political institutions, in consequence, differed in several important respects from those of the continent. There was a single parliament for the entire country; no regional estates as in France, Spain, and the Netherlands. There was also no divisions into towns, clergy, and nobility. But the more centralized feudal structure did not gainsay that the crown could not overstep the traditional liberties of the barons as the Magna Carta attests.

We can now turn to examining the evolving bargaining strength (and the three underlying determinants) of ruler versus constituent that shaped the belief structure and the path of each polity. Take the Netherlands. The productive town economies stood to gain substantially by the political order and protection of property rights provided by the Burgundians and then by Charles V. The structure of the economy built around export trades provided the means for easy-to-collect taxes on trade but not at a level to adversely affect the comparative advantage of those export trades. The liberty to come and go, buy and sell as they saw fit led to the evolution of efficient economic markets. But when Philip II altered the "contractual agreement" the Seven Provinces became convinced that they could only prosper with independence. The resistance was initiated by the States General which in 1581 issued the Act of Abjuration of allegiance to Philip II and claimed sovereignty for the Provinces themselves. The powers of the newly independent country resided with each province (which voted as a unit) and a unanimity rule meant that the States General could only act with the unanimous approval of the Seven Provinces.

Cumbersome as that process was, this political structure survived. The polity not only evolved the elements of political representation and democratic decision rules but equally supported religious toleration. The belief structure that had evolved to shape the independent polity was more pragmatic than "intellectual," a consequence of the incremental evolution of the bargaining strength of constituents and rulers.

As with the Netherlands it was England's external trade that provided an increasing share of crown revenue with taxes on wine, general merchandise and wool cloth; but it was the wool export trade that was the backbone of augmented crown revenue. Eileen Power's classic story of the wool trade (1941) describes the exchange between the three groups involved in that trade: the wool growers as represented in Parliament, the merchants of the staple, and the Crown. The merchants achieved a monopoly of the export trade and a depot in Calais, Parliament received the right to set the tax, and the crown received the revenue. Stubbs (1896, 3:599) summarized the exchange as follows: "The admission of the right of parliament to legislate, to enquire into abuses, and to share in the guidance of national policy, was practically purchased by the money granted to Edward I and Edward III..."

With the Tudors the English crown was at the zenith of its power but it never achieved the unilateral control over taxing power that the crowns of France and Spain achieved. The confiscation of monastery lands and possessions by Henry VIII alienated many peers and much of the clergy and as a consequence "Henry had need of the House of Commons and he cultivated it with sedulous care" (Elton, 1953, 4). The Stuarts inherited what the Tudors had sown and the evolving controversy between the Crown and Parliament is a well known tale. Two aspects of this controversy are noteworthy for this analysis. One was the evolving perception of the common law as the supreme law of the land--a position notably championed by Sir Edward Coke--and the other was the connection made between monopoly and a denial of liberty as embodied in the Crown grants of monopoly privileges.

IV

A fundamental revolution occurred in the second half of the nineteenth century which I have termed the second economic revolution.⁷ This revolution was a consequence of a change in the stock of knowledge arising from the development and implementation of scientific disciplines. It resulted in the systematic wedding of science and technology and a basic transformation in the organization and structure of production and distribution (see Chandler, 1977). The overall implications for economies that could take advantage of this technology were increasing returns and consequent high rates of economic growth--characteristics of the western economies for the past century and a half. But taking advantage of this technology entailed a wholesale reorganization of economies to realize that potential. In those western economies that have, at least partially, realized this potential the result has been stresses and strains that have threatened and do threaten their

⁷. The first economic revolution was the development of agriculture in the eighth millenium B.C. in the Fertile Crescent

continued adaptive efficiency. For the rest of the world the inability to reorganize has prevented them from realizing this productive potential and produced "underdevelopment" and political instability. It is an extraordinary irony that Karl Marx, who first pointed out the necessity of restructuring societies in order to realize the potential of a new technology, should have been responsible for the creation of economies that have foundered on this precise issue. Let me first examine the micro level characteristics of the organizational requirements before turning to the macro level societal implications.

Realizing the gains from a world of specialization requires occupational and territorial specialization on an unprecedented scale and in consequence the number of exchanges grows exponentially. In order to realize the gains from the productive potential associated with a technology of increasing returns one has to invest enormous resources in transacting. In the United States, for example, the labor force grew from 29 million to 80 million between 1900 and 1970; during that period production workers grew from 10 million to 29 million, while white collar workers (the great majority of whom are engaged in transacting) increased from 5 million to 38 million. The transaction sector (that part of transaction costs that goes through the market and therefore can be measured) in the United States in 1970 made up 45 percent of GNP (Wallis and North, 1986).

Let me briefly elaborate some of the measurement and enforcement problems that account for the size of the transaction sector. Necessary to be able to realize the gains of a world of specialization are control over quality in the lengthening production chain and a solution to the problems of increasingly costly principal/agent relationships. Much technology indeed is designed to reduce transaction costs by substituting capital for labor or by reducing the degrees of freedom of the worker in the production process and by automatically measuring the quality of intermediate goods. An underlying problem is that of measuring inputs and outputs so that one can ascertain the contribution of individual factors and the output at successive stages of production. For inputs there is no agreed upon measure of the contribution of an individual input. Equally there is room for conflict over the consequent payment to factors of production. For output, not only is there residual unpriced output, that is waste and pollutants, but also there are complicated costs of specifying the desired properties of the goods and services produced at each stage in the production process.

Another characteristic of this new technology is that firms have large fixed capital investments with a long life and (frequently) low alternative scrap value. As a result the exchange process embodied in contracts has to be extended over long periods of time, which entails uncertainty about prices and costs and the possibility of opportunistic behavior on the part of one of the parties to the exchange. A number of organizational problems emerge from these characteristics associated with this technology.

First, increased resources are necessary to measure the quality of output. Sorting, grading, labeling, trade marks, warranties and licensing are all, albeit costly and imperfect, devices to measure the characteristics of goods and services. Despite the existence of such devices the dissipation of income is evident all around us in the difficulty of measuring the quality of automobile repairs, in evaluating the safety characteristics of products and the quality of medical services, or in measuring educational output.

Second, while team production permits economies of scale to be realized, it does so at the cost of worker alienation and shirking. The "discipline" of the factory is no more

than a response to the control problem of shirking in team production. From the perspective of the employer the discipline consists of rules, regulations, incentives, and punishments essential to effective performance. Innovations such as time and motion studies are methods of measuring individual performance. From the viewpoint of the worker they are inhuman devices to foster speedups and exploitation. Since there is no agreed upon measure of output that constitutes contract performance, both are right.

Third, the potential gains from opportunistic behavior increase and lead to strategic behavior both within the firm (labor-employer relations, for example) and in contractual behavior between firms. Everywhere in factor and product markets the gains from withholding services or altering the terms of agreement at strategic points offer large potential gains.

Fourth, the development of large scale hierarchies produces the familiar problems of bureaucracy. The multiplication of rules and regulations inside large organizations to control shirking and principal/agent problems results in rigidities, income dissipation, and the loss of flexibility essential to adaptive efficiency.

Finally there are external effects: the unpriced costs reflected in the modern environmental crisis. The interdependence of a world of specialization and division of labor increases exponentially the imposition of costs on third parties.

The institutional and organizational restructuring necessary to take advantage of this technology are, however, much more fundamental than restructuring economic organization--although that task, the creation of efficient markets, is complicated enough. The entire structure of society must be transformed. This technology and accompanying scale economies entails specialization, minute division of labor, impersonal exchange and urban societies. Uprooted are all the old informal constraints built around the family, personal relationships, and repetitive individual exchanges. Indeed the basic traditional functions of the family: education, employment (the family enterprise), and insurance are either eliminated or severely circumscribed. New formal rules and organizations and an increased role of government replace them.

The contention of Marxists was that these problems were a consequence of capitalism and that the inherent contradictions between the new technology and the consequent organization of capitalism would lead to its demise. The Marxists were wrong that the problems were a consequence of capitalism; they are ubiquitous to any society that attempts to adopt the technology of the second economic revolution. However, as the foregoing paragraphs have attempted to make clear, Marxists were right in viewing the tension arising between the new technology and organization as a fundamental dilemma. These tensions have only partially been resolved in the market economies of the western world. The growth of government, the disintegration of the family, the incentive incompatibility of many modern political and economic hierarchical organizations are all symptoms of the consequent problems besetting western economies.

However, it has been the relative flexibility of the institutions of the western world--both economic and political--that has been the mitigating factor in dealing with these problems. Adaptive efficiency, while far from perfect in the western world, accounts for the degree of success that such institutions have experienced. The basic institutional framework has encouraged the development of political and economic organizations that have replaced (however imperfectly) the traditional functions of the family; mitigated the

insecurity associated with a world of specialization; evolved flexible economic organization that has induced low cost transacting; resolved some of the incentive incompatibilities of hierarchies, and encouraged creative entrepreneurial talent; and tackled (again very imperfectly) the external effects that are not only environmental but also social in an urban world.

