Transaction Costs, Trust,
and the Structuring of Markets


Full text available as: PDF - Requires Adobe Acrobat Reader or other PDF viewer

Abstract

This paper examines the institutional arrangements that develop when the risks of opportunism and other contributors to transaction costs are high but transactions are nevertheless necessary for economic efficiency.

Williamson’s argument that high levels of transaction costs lead firms to choose vertical integration (hierarchical organisation) in preference to using markets is well known. But this is far from the whole story because markets are not uniform. In fact, modern “market economies” embrace a wide range of institutions for exchanging ownership of goods and services. The types of transactions undertaken are not only varied, but some important markets are themselves hierarchies that are structured in ways that are analogous to firms precisely in order to reduce their costs of operation, including the transaction costs that arise from using them.

In this paper, I look at the evolution of a select but important group of markets that have been consciously constructed over long periods and with frequent modifications because of environmental change and learning by participants. These are markets in which the use of up-to-date information is especially important because conditions may alter quickly, and in which risk, uncertainty, and the potential for opportunistic behaviour are factors that affect their operations in significant ways. When transaction and agency costs arise in such markets, responses have concentrated on finding mechanisms for reducing them to tolerable levels rather than on abandoning transactions altogether through the internalisation of activities. The logic of constructed markets is illustrated by an examination of the evolution of membership regulations by the London and New York Stock Exchanges in the nineteenth and early twentieth centuries.

EPrint Type: Departmental Technical Report
Keywords: Transaction costs, trust, constructed markets, stock markets
Subjects: 340000 Economics; 340205 Industry Economics and Industrial Organisation
ID Code: JEL Classification L52 (Transactional Relationships); N20 (History of Financial Institutions); and B52 (heterodox methodologies - Evolutionary).

Deposited By: Weaver, Belinda

Paul L. Robertson
School of Economics
University of Queensland and
The Australian Innovation Research Centre,
University of Tasmania
plr090634-at-hotmail.com
TRANSACTION COSTS, TRUST, AND THE STRUCTURING OF MARKETS

Paul L. Robertson
University of Tasmania/University of Queensland
plr090634-at-hotmail.com

Introduction

The boundaries of firms and markets are inextricably related. The exact location of the boundaries of firms, that delineate the activities undertaken internally, are often attributed to “market failure” – to inefficiencies and other “frictions” that make markets expensive to use and therefore induce firms to produce goods and services internally that they might otherwise purchase in markets. In this formulation, markets are actually given the starring role even though attention usually focuses on firms because economists tend to view market transactions as more natural in some sense than the internalization of activities that might otherwise be purchased. In his seminal 1937 article on transaction costs, Ronald Coase offers a quotation from Sir Arthur Salter as a typical economist’s view of the way in which market economies function (Coase, 1937; 387):¹

The normal economic system works itself. For its current operation it is under no central control, it needs no central survey. Over the whole range of human activity and human need, supply is adjusted to demand, and production to consumption by a process that is automatic, elastic and responsive.

In other words, the price system operating through markets provides all of the coordination necessary for an economy to operate as near to optimally as makes no practical difference. While firms may benefit through narrow specialization in a Smithian sense, they are unlikely to benefit from vertical integration as long as markets are frictionless (that is, costless to use). Moreover, adding activities may be increase production costs since it is hard to be a master of all trades. As a result, transaction cost economics explains the existence of vertically-integrated firms with complicated organizational structures by claiming that market failure may drive firms to broaden their scope of activities.

The converse of the argument is that the presence of markets with relatively low transaction costs will discourage vertical integration. Furthermore, the extent of vertical integration can be expected to vary as transaction costs rise and fall (Langlois, 1992). Therefore, at least in this formulation, both the boundaries of firms (the range of activities that they undertake) and the boundaries of markets (the range of activities undertaken and the cast of players) are primarily determined by the costs of using markets rather than by the productive efficiency of firms.

But this is far from the whole story because markets are far from uniform. In fact, modern “market economies” embrace a wide range of institutions for exchanging ownership of goods and services. The types of transactions undertaken are not only varied, but some important markets are also hierarchies that are structured in ways that are analogous to firms precisely

¹ Coase (1993a [1988]) also mentions Salter’s passage in his fiftieth anniversary account of the origins of his article.
in order to reduce their costs of operation, including the transaction costs that arise from using them.

In this paper, I look at the evolution of a select but important group of markets that have been consciously constructed, albeit over long periods and with frequent modifications because of environmental change and learning by participants. These include stock exchanges, insurance exchanges such as Lloyd’s, and various markets involving shipping and world trade, markets in which the use of up-to-date information is especially important because conditions may alter quickly and in which risk, uncertainty, and the potential for opportunistic behavior are factors that affect their operations in significant ways. Their productivity as markets is (or historically has been) so high that their replacement by hierarchies is virtually unthinkable because they allow for exchanges that could not otherwise be accomplished smoothly. As a result, when transaction and agency costs arise in such markets, responses have concentrated on finding mechanisms for reducing them to tolerable levels rather than on abandoning transactions altogether through the internalization of activities.

In the first section, I try to define what a market is, a problem that turns out to be surprisingly elusive. The second section investigates the sorts of factors that may affect the smooth operations of the types of markets under consideration by examining both production costs and transactions costs of various types. In the third section, the analysis is applied to the development of stock exchanges since the seventeenth century to demonstrate how they have been constructed with relatively confined boundaries in order to restrict transaction and agency costs, improve their internal efficiency, and provide ways of coping with inherent uncertainty as well as risk. The concluding section applies some observations made in the earlier discussion to broader questions concerning transaction costs and the boundaries of firms and markets.

1. Markets and The Market

Although “markets” lie at the core of modern economies as well as underpinning neoclassical economic theory, in practice they have such a wide range of characteristics that, to use a biological analogy, in their level of diversity they are closer to a genus, or even a family or an order, than to a species.

In its most general sense, a market (or The Market) is an ideal type, a consciously chosen selection of characteristics that have been logically configured by economists to examine how the price system functions. Although perfectly competitive markets lie at the heart of the analysis, variations based on more limited forms of competition (monopoly, duopoly, oligopoly, etc.) have also been dissected and evaluated by economists. The importance of The Market is not its realism since it is doubtful if perfect competition has often prevailed in the real world, but that the analysis provides a series of benchmarks against which the performance of other market structures may be judged. For example, Caves (1964; 15) contends that “Market structure is important because it determines the behavior of firms in the industry, and that behavior in turn determines the quality of the industry’s performance.” On examination, this leads to the conclusion that highly competitive markets with high levels

---

2 Economists are, of course, not the only people who have been concerned with defining markets. Max Weber, for instance, gave considerable attention to the issue, in the end deciding that markets have “an amorphous structure” (quoted in Swedberg (2000), p. 379).
of performance are superior, a deduction that Caves (1992; 15) defines as “our normative appraisal of the social quality of the allocation of resources that results from a market’s conduct”. As I discuss below, this analysis of The Market has contributed not only to a preference for high levels of competition, but also to the standards of behavior and performance that groups have tried to generate when constructing markets in order to reduce market failure.

This approach to markets has tended to squash consideration of markets and their alternatives in some areas where we might expect a strict and workable definition of markets to be essential. In a survey of microeconomics textbooks, for example, I found that markets are often not defined at all or, if they are, that the definitions provide little for students to ponder. In this, they are following Williamson (1975, 20) in contending that “in the beginning there were markets”. While Williamson notes that he is taking this position for “for expositional convenience”, however, some other authors do not seem to recognize that there is an issue at stake. Expositional convenience may well justify some cutting of corners, and it is clear that in many areas in modern capitalist economies the existence of markets is so common as to virtually undebatable, but recent events in China, Eastern Europe, and parts of the former Soviet Union have demonstrated that the preconditions for the efficient operation of markets are not universally available and that conscious intervention may be necessary to build a market economy. Laws securing property rights, for example, are constructs and by no means “natural” in the minds of people in many societies. (North, 2005; Greif, 2006)

Popular dictionaries of economics are not much better at capturing the meaning of “market”.3 The *Oxford Dictionary of Economics* (Black, 2002; 288) offers as a first definition that a market is “A place or institution in which buyers and sellers of a good or asset meet.” It asserts that “Markets facilitate trade in goods…; in securities…; in labour services…; or in foreign exchange…”, but does not discuss how this facilitation occurs. The second definition in the *Oxford Dictionary* refers to a market economy and in essence echoes the views of Caves on the benefits flowing from high degrees of competition. This definition notes that “An efficient market is one where prices reflect all available information about the good or asset concerned (289)” and lists a number of causes of market failure, but gives no indication of how market failure might be avoided and efficiency achieved. The *Collins Dictionary of Economics* (Pass et al., 2005; 324) says that a market is “An exchange mechanism that brings together buyers and sellers of a PRODUCT, FACTOR OF PRODUCTION or FINANCIAL SECURITY (see TRANSACTION).” The definition notes a number of characteristics of markets and describes various types of competition, but (like the *Oxford Dictionary*) does not discuss how markets might operate in practice to achieve efficiency and high levels of competition.

The correct conclusion to be drawn from these definitions, which is not mentioned explicitly in any of the accounts, is that markets are by no means homogeneous entities. This is interesting because, while it does not undermine neo-classical analysis on the advantages of competition and market efficiency, this variety of forms leads to a distinctly fuzzy picture of what goes on in a market economy. The most obvious consequence is in the area of competition policy because fuzziness seems to imply that, depending on circumstances, a number of types of markets might achieve acceptable levels of social welfare. Similarly, by not specifying what happens in a market beyond the basic statement that it is a place in which buyers and sellers exchange assets, such vague definitions beg the question of what sorts of

---

3 For a thought-provoking and controversial discussion of the meaning of “market” and “marketplace” from a sociological standpoint, see Callon (1998).
self-regulation and other aspects of constructing and limiting the operations of markets might be considered to be appropriate.

Although economists may refer to The Market and the market economy, in practice there is a huge number of markets in a developed economy. Many of these markets do not have a distinct institutional or physical presence: Lemonade can be bought at a supermarket, a restaurant, or from a child sitting in front of her house. Nor is it necessarily true that a homogeneous product will be sold in a distinct market: The same airline ticket may be purchased from the airline, a travel agent, or online as part of a package tour. What, then, determines the boundaries of “the market” for lemonade or airline tickets? Is it even meaningful to speak of markets in such miscellaneous circumstances?

Of more interest here are specialized markets which are often sharply defined both institutionally and physically. These include markets for particular classes of good and services such as securities, marine insurance, and a wide variety of commodities. Some of these markets are centuries old, but others (markets for derivatives, for example) are of more recent origin. Moreover, the category has a realistic chance of growing in the future because various e-markets have many similar characteristics even though their operations are as geographically diffuse as conditions allow.

The distinguishing feature of these markets is not that they are specialized in terms of the items traded (some sell more than one type of product) but that their operation is restricted to people who are willing to subscribe to a formal set of rules. In general, as is discussed in more detail in the next section, rules are needed to restrict transaction costs in cases in which internalization is not a plausible alternative to operating through a market. Lloyd’s, for example, describers itself as “the world’s leading insurance market providing specialist insurance services” (Lloyd’s, 2006). It was established to spread risk widely by issuing policies backed by large pools (syndicates) of people (Names) with funds to invest. Since the backers have to be reliable and fraud needs to be kept within tight limits to allow a reasonable prospect of profitability from the sale of marine and other types of insurance, the insurers try to block potential opportunism by vetting Names before they are allowed to invest, just as it assesses the risks facing shipowners and others who want to buy policies. While internalization in the form of self-insurance would be a possible alternative for potential policy holders who are worried about being cheated when they make legitimate claims, many people or firms would find the risk excessive. As a result, they are willing to pay slightly more in order to do business with insurers who operate in a formally demarcated and self-regulated market. Equally importantly, the closed nature of Lloyd’s protects the investors from serious risk of default on the part of other names in their syndicates.

In effect, this category comprises markets that have been constructed quite consciously to provide the benefits that economists have dictated should exist in The Market but which are frequently missing from markets in the everyday world.4 In practice, markets do not have to offer free access to all comers to be highly competitive. All that is required is a sufficient number of participants to allow prices to be set efficiently through competitive bidding. The

4 This is not to imply that the people who established and later fined tuned the operations of specialized models consciously used economic theory as a guide, especially since a number of the markets were begun before economists had codified the benefits of high levels of competition. As practitioners, however, they were aware of their practical needs as traders, needs that were eventually codified by generations of microeconomists. The point is discussed further in Section 4 in relation to arguments by Callon (1998) and MacKenzie and Millo (2003).
main focus of the regulations that these markets laid down and enforced was on the frictions that are now termed transaction costs. In the process, these regulations generated other types of transaction costs, but these are of a type that is often overlooked – barriers that enabled markets to function more smoothly than they would have if unregulated. The upshot has been that the “production costs” involved in using these markets (that is, in gaining the benefits that flow from transacting in markets rather than through other types of trading arrangements) have in general been reduced by managing transaction costs in a systematic way.

2. Why Construct Markets?

The factors that potentially influence the design of constructed markets are contingent on the institutional and physical environment in which each market operates and on the nature of the goods or services being traded. They can be divided into the traditional categories of production and transaction costs and, following Langlois (forthcoming), transaction costs can be further subdivided into fixed costs, costs that are a function of time, and costs that are a function of the number of exchanges or volume of trade undertaken (Figure 1). My argument is that, for constructed markets, many of the costs that are generally designated as transaction costs are actually production costs and that, far from being frictions, they contribute substantially to the efficiency of the markets in question.

The principal advantage that conscious structuring confers on markets is that it allows all of these types of costs to be managed. This does not mean that the costs can be eliminated but that, under favorable circumstances, they may be reduced in various ways that conduce to greater efficiency in the operations of the markets. In Williamson’s famous dichotomy (1975, 1985), exchanges may be dealt with through either markets or hierarchies. Other intermediate forms have since been promoted including strategic ventures and other forms of cooperation (Thompson, et al., 1991). Coase (1960) has suggested (although hardly endorsed) the proposition that government regulation could be used to reduce transaction costs. The alternative discussed in this paper is the self-regulation of markets that are, in most cases, owned and managed by their participants. Because they are deliberately limited in both their functions and in the number and characteristics of their players, constructed markets have an ability to shape their environments and outcomes that is not open to markets as traditionally conceived.

As Coase (1993b [1988]) has emphasized, neither production costs nor transaction costs should be analyzed in isolation because it is their sum that provides the basis for deciding whether to use markets, hierarchies, or other organizational forms. The main purpose of a firm is to run a business – to produce goods or services for sale at a profit (Coase, 1993b [1988]); Langlois and Robertson, 1995). Similarly, the main purpose of a market is to allow transactions to be undertaken. Thus markets should be thought of as also having production costs unless one assumes that transactions somehow occur through some supernatural mechanism that does not require human effort or other inputs. In common with firms, markets may face a variety of costs, not all of which can be outsourced. For instance, a market may need a venue or marketplace (essentially the first definition of a market offered up in the Oxford Dictionary of Economics (2002)). In some cases, this cost may be borne in a decentralized fashion, as when the market for new automobiles is spread around a number of individually-owned dealerships. In others, the cost can be shoved off onto others, as in the case of the “curbside exchange” that, before it evolved into the American Stock Exchange, met in the streets of New York. Likewise, the personnel needed to maintain a market may
<table>
<thead>
<tr>
<th>Fixed Costs</th>
<th>Costs that are Function of Time</th>
<th>Costs that are a function of the number of exchanges or volume of trade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examples:</strong> Legal, organization, and technological standards; hostages and bonds; locks, closed circuit TV</td>
<td><strong>Examples:</strong> Salaries of police, supervisors and other monitors; monthly protection money; maintenance of fixed investments</td>
<td><strong>Examples:</strong> Brokerage fees, commissions; insurance premia; queuing at the bank, ATM fees; inspection and regulatory fees; per-transaction bribe</td>
</tr>
</tbody>
</table>

**Costs of Protecting Property Rights**

**Costs of Protecting Property Rights**

**Neoclassical Transaction Costs**

Source: Based on Langlois (forthcoming), Figure 1.

**Figure 1**

Types of (Transaction) Costs
be deployed and paid for in a decentralized manner, or they may be employees of the market, or both. Finally, markets may need an initial fund of capital to get them started and keep them operating smoothly. This capital could be subscribed by participants in the market and administered centrally, or it might be provided and administered by individual participants who, in effect, barter their money and services for providing a lubricant for part of a market’s operations for similar services provided by other participants. The choice of a structure to administer, and distribute the costs of production differs from market to market depending on a host of factors such as the number of participants. Neither very large nor very small numbers, for example, could be expected to maintain distinct premises. Constructed markets can adopt rules and procedures that allow them to undertake these tasks efficiently while still maintaining their essential function of facilitating transactions.

A second possible reason for the construction of markets is the reduction in transaction costs that may derive from specialization. By trading in markets with limited scope, the types of efficiencies that Smith (1937 [1776]) noted in the early chapters of The Wealth of Nations may become available: Greater mental and physical adeptness, greater innovativeness, and better use of time. Because of their technical nature and significant reliance on up-to-date information, specialization can certainly improve the efficiency of many of these markets, but the matter is not straightforward. As a result, the scope of markets that are seemingly dealing in the same goods may differ substantially. In part, this is because of economies of scope since hard-to-acquire technical knowledge that is relevant to one field may also apply to others, encouraging people with rare skills to diversify their activities if they have spare resources. In other words, the division of labor is limited by the extent of the market. But there may be more involved as there is no reason to think that there is always a complete correspondence between the division of labor among markets and the degree of specialization of the people who use those markets. A market that is substantial enough to be highly specialized in terms of the products traded may nevertheless absorb only a fraction of the time of the buyers and sellers who participate. As I show below, for instance, in their early days stock exchanges often handled only a proportion of the activities of many of their members, who were as a result obliged to “saunter” in the course of their work. Thus it is reasonable to expect that the costs of sauntering would be balanced against the advantages that arise from market specialization.

An additional justification for limiting numbers, one that economists are guaranteed to mention, is that structured and regulated markets may be used to drive up profits by curtailing competition. This obviously is a possibility in the specialized markets under discussion here, but there are limits to how severely competition can be limited when the purpose of the market is to generate conditions that approximate those of The Market. If an efficient market is one in which all participants have access to prices that convey “all available information about the good or asset concerned” (Black, 2002, 289; Bernstein, 1992; Lo, 1997), then there must be enough buyers and sellers to insure that the prices are determined through a process involving a fairly high degree of competition. On the other hand, bounded rationality must be restricted as far as possible because all participants need to be fully up-to-date on prices and the underlying information that they reflect. In practice, this may place limits on how

---

5 Near the end of the seventeenth century, for instance, the City of London licensed “brokers” but did not stipulate what products each was to deal in (Morgan and Thomas, 1962).

6 A lack of correspondence in the degrees of specialization and integration of different factors of production may affect the division of labor in other respects. For example, over the past two centuries the integrative capacity of machinery to perform many tasks has had serious effects on the scope of activities of individual workers (Ames and Rosenberg, 1965; Robertson and Alston, 1992).
large the markets can be. If it is not overdone, therefore, limiting the size of markets can increase their efficiency.

The prospect of competition is reinforced by the possible existence of other markets whose members trade in the same class of assets. In a number of cases, as in insurance and securities, multiple markets have been in competition, offering lower charges or additional services to attract customers.

Limitations on size can also be necessary to keep the level of transaction costs (Figure 1) within bounds. But, as has already been noted, some of the most often cited types of transaction costs are more accurately regarded as production costs. In real markets certain kinds of behavior that generate transaction costs are necessary because, in their absence, there can be no transactions (Campbell and Harris, 1993; Macneil, 1981 and 1982). These costs result from what Nooteboom (2002) calls “enabling constraints”. As Campbell and Harris (1993, 178) put it,

The negotiating, information gathering, organizing, and so on within which transactions take place are not only costs, they are also the social relations which are essentially facilitative of the transaction. All actions, including all transactions, can take place only within a constitutive social system. If one really took away all the costs of exchanging, the exchange would not take place cost-free. It simply would not take place at all (emphases in original).

Macneil (1981, 1022) goes further by contending that,

Because it is impossible to conduct exchange without transaction costs, and since they are variable, they are as much a factor of production as are capital and labor.\(^7\)

Together, these comments justify Arrow’s conclusion\(^8\) that transaction costs are the “costs of running the economic system”. As all firms in a capitalist economy sooner or later need to confront markets, internalizing transactions can only be a postponement of the inevitable.\(^9\) As an alternative to internalization of activities, firms may adopt a program of reducing transaction costs, for example by encouraging the establishment of several suppliers of a particular input to reduce the likelihood of hold-up. Although neoclassical economists have tended to ignore transaction costs, from the point of view of a firm these are not particularly mysterious and are more likely to be thought of a normal cost of doing business.

The correspondence between production and transaction costs is clear for constructed markets whose product is transactions. The managers of these markets have great incentive to reduce transaction costs by regulating and monitoring (i.e. managing) both \textit{ex ante} and \textit{ex post} activity. As far as possible, they need to reduce the impact of bounded rationality in order to increase market efficiency, but also to reduce opportunistic activity and to limit risk.

\(^7\) Baldwin and Clark have referred to these as “mundane transaction costs”, but their importance is defended in Jacobides (2005) and Langlois (forthcoming).

\(^8\) As quoted in Macneil (1981, 1022).

\(^9\) Which is not to imply that internalization is useless. Among other things, it may allow a firm to select which markets in which it does business both as a buyer or a seller, permitting it to choose ones in which transaction costs are relatively modest (and competitive conditions relatively favorable, as in Porter’s (1980) five forces model).
and curtail the effects of uncertainty as markets will not function adequately if the risk of default is excessive.\(^\text{10}\)

In the past quarter century, opportunism has occupied a central place in discussions of transaction costs (Williamson, 1985). Because of asymmetric distributions of knowledge, some market participants may be able to turn transactions to their advantage if they are willing to engage in “self-interest seeking with guile” (Williamson, 1993, 458). One way of dealing with this, which is emphasized by Williamson (1975 and 1985), is to withdraw from markets and keep activities in-house where they can be supervised more closely, but in other cases it is possible to reduce opportunism to levels that still allow transactions to take place profitably.\(^\text{11}\) As eliminating transaction costs also incurs costs (Langlois, forthcoming), it is again necessary to ensure that the best balance of total costs is found.

One way of reducing opportunism is to deal only with people who are regarded as unlikely to act with guile. Unfortunately, such people are not always easy to identify, especially since past behavior, even if known, may not be a good indicator of future behavior, which is what parties to a transaction really need to know. Two broad strategies, neither of which is foolproof, can be followed in attempts to avoid having to do business with bad trading partners. The first is to achieve an alignment of interests such that non-opportunistic behavior by both parties will lead to win-win outcomes while opportunism on the part of either party generates lose-lose outcomes. Transaction costs are therefore reduced since all participants know that both they and other parties will avoid opportunistic activities for fear of being penalized while non-opportunistic behavior leads to rewards all round. An alignment of interests can be achieved in a number of ways, including the use of formal or informal contracts that specify performance levels and their associated positive and negative payoffs.

As contracts can be expensive to draw up and enforce, however, other means of aligning interests may be superior. One alternative means is to deal as far as possible with others whose success is to some extent directly related to one’s own, for example family members. When a substantial proportion of assets are held jointly, it is in the interests of each party that others should thrive.\(^\text{12}\) But reliance on relatives can be excessively confining. Not only is there no guarantee that a relative with the proper assets can be found to take the other side of a transaction, but, no matter how well aligned their interests, family members may not be as skilled or intelligent as outsiders, and the sum of a family’s assets may be inadequate for some deals.\(^\text{13}\)

As a result, other sorting devices may be needed to reduce the risk of opportunism. This brings up questions of trust, an issue that has received uneven attention. Ordinary usage suggests that “trust” comes in a number of varieties and degrees of strength. Thus we might trust someone with our lives whom we would not trust to make a good cup of coffee.

\(^{10}\) Better flows of knowledge and information can provide both \textit{ex ante} and \textit{ex post} remedies to opportunism, but other \textit{ex ante} controls may also be valuable in eliminating opportunism and thus reducing the need for, and the costs of, monitoring transactions (Lindenberg, 2000).

\(^{11}\) That is, to a level at which transaction costs are smaller than the gains that arise from using markets rather than alternative arrangements.

\(^{12}\) Many of the same attributes are relevant in determining patterns of organization in family firms (Colli, 2003), but business relationships among relatives can also be established on an \textit{ad hoc} basis.

\(^{13}\) Moreover, as Lindenberg (2000) points out, emotional blackmail and other tactics can lead to profits and losses being distributed in ways that do not correspond to the original amounts contributed or to the distributions that would be expected when dealing with outsiders.
Williamson (1993) has introduced another criterion by claiming that trust should not be calculative but rather equated with “blind trust”, that is with believing that another person will behave in our best interests solely on the basic of some heuristic (motherly love, perhaps) rather than from a strict and thorough examination of the person’s background, capabilities, or other characteristics that are relevant to their motives and ultimate performance. Once we exhibit calculativeness by consciously weighing up the factors involved in another’s performance and assessing the risks involved in a relationship, Williamson claims that trust is no longer involved.14 Other commentators (Nooteboom, 2000 and 2002; Lindenberg, 2000) have contested Williamson’s argument and considerably illuminated the issues in question. Here, I accept the thrust of Williamson’s point that calculativeness is often involved in assessing relationships with others but deny that this undermines trust as a way of reducing transaction costs. Indeed, Williamson’s claim sidesteps the main point, which is to determine how information can be used to reduce transaction costs. Seabright (2004, 64) captures the right tone when he claims that,

when I say I can trust a stranger, I do not mean that I like him, have any curiosity about him as a person, or care in any deep sense about what happens to him. The point is that I do not need to like or care about him in order to be able to deal confidently and reliably with him.15

Naturally, if we know that people with whom we want to do business will not cheat us because they are pure of heart, then transaction costs will be reduced because there is no longer a risk of opportunism.16 But how can we confidently decide if a person can be dealt with reliably? A number of techniques may be used, although none totally eliminates risk in the face of inherent uncertainty.

One way of proceeding, which is known in many societies,17 is to deal as far as possible with the members of one’s peer groups, for instance with people from the same locality, school or regiment. This is a way of economizing on information in selecting colleagues and trading partners as well as of reducing the incidence of opportunism. Bad hats, as well as people who succumb to short-term temptation, can be found everywhere, but they may be easier to identify among a group whom the selector is already familiar with and whose mores he understands. Furthermore, Lindenberg (2000) argues that allegiance to a group leads to greater stability of normative frames in which people are eager to “act appropriately” or “to do the right thing” than is likely to occur in dyadic relationships. In the case of a constructed market, the behavior of participants may be constrained by membership in two groups, the larger social category from which they are drawn and their immediate colleagues in the market itself. Whether from sensibility or fear of sanctions, parties to a transaction within a

---

14 At this point, Williamson’s argument reduces to a semantic quibble. Some people might prefer, for example, to say that blind trust is not trust at all but “faith”.

15 This view of trust closely parallels Weber’s formulation of the sort of social action that takes place when transacting in a market environment (Swedberg, 2000).

16 As there are other types of transaction costs than opportunism (Figure 1), it would still not be costless to use markets. Still, if we somehow know who is and is not well disposed towards us and likely to pose a threat, this would not only reduce the probable loss that might arise directly from being cheated but also the costs associated with monitoring for opportunism and perhaps some of the costs involving in checking out potential colleagues before entering into a transaction.

17 For China, see Redding’s work (1990). Lie (1998) claims that South Korea has been aiming for national homogenization but this has not totally discouraged reliance on peer groups at more micro levels. See also Ungson, et al. (1997).
well-defined and easily monitored group are likely to be reluctant to “let the side down” as well as to shame themselves.

Opportunism is not the only reason for choosing trading partners with great care. Unpredictable shocks can cause disaster in a world of uncertainty, even shocks that apply only indirectly to an associate. When there is unlimited liability for debts – which was the norm when many markets were constructed and still applies to transactions in many financial markets\(^{18}\) – a person can be forced to pay debts that might more appropriately be assigned to his partners. Thus there is a potential for asymmetry in returns as an investor can bear risk for his partners without being entitled to share fully in their gains. Following recent disasters, the ability of Lloyd’s syndicates (of which there are presently 63) to vet the finances of their participating Names has kept claims from becoming the responsibility of only a few of the more solvent Names and assured that the claims would be paid by the syndicates as originally contracted. This has caused extreme discomfort for some of the less wealthy Names but has also confined the damage within Lloyd’s and ensured that it could continue to issue policies. The pitfalls of partnerships with unlimited liability extend even further. When a risk may arise not just from the indiscretions or misjudgments of a partner but from those of that person’s other partners, a default can lead to a disaster spreading several removes from its origins. This again supports restricting the operations of a market to people from a group that can be easily monitored as it is necessary to collect information that extends beyond things that, at first glance, seem directly relevant to potential participants in the market.

The benefits of self-regulation in constructed markets are therefore manifold. Constructed markets can decide their scope of activities in order to achieve appropriate levels of specialization. Within the normal meaning of production costs, these markets are able to control their own venues and some proportion of their personnel and capital. They are also able to regulate to ensure that their markets are relatively efficient and transparent by placing controls on the number of participants. Most importantly, because they can control the number of participants, they can also scrutinize their characteristics and monitor their activities (Lindenberg, 2000) relatively cheaply. As a result, constructed markets are able to deal with substantial levels of opportunism and risk without impairing their ability to function efficiently.

3. Stock Exchanges as Constructed Markets

The operations of constructed markets can be illustrated by examining aspects of the history of the two most important stock exchanges, in London and New York.\(^{19}\) While their stories run parallel in many respects, they also show the effects of local environments on institutional development.

Institutional economists and economic historians have already charted the contingent nature of the establishment and increasing sophistication of markets over the past thousand years and they have also examined self-regulation as a device for reducing transaction costs as far

---

\(^{18}\) More recently, firms operating in a market may still have unlimited liability for debts even though they are incorporated and the debt of their owners is limited. Managers are understandably reluctant to allow their firms to go bankrupt despite the boundaries on damage suffered by the firms’ owners.

\(^{19}\) For more complete accounts of the development of the London and New York Stock Exchanges, see Michie (1987, 1988, and 2001), Morgan and Thomas (1962), Sobel (1964), and Geisst (2004). These exchanges did not provide the only model for development. As Weber (2000 [1894]) discusses, their high degree of self-regulation contrasted sharply with the greater degree of government regulation common on the Continent.
back as the Middle Ages (North, 2005; Greif, 2006; Milgrom, North and Weingast, 1990; Greif, Milgrom, and Weingast, 1994). The evolution of stock exchanges provides a different perspective, however, because of their role in developing modern institutions and their continuing importance. They therefore offer a set of good examples of how specialized markets have been consciously constructed over an extended period on a trial-and-error basis in order to cope with changing circumstances. This can be demonstrated by examining the early development of the London (LSE) and New York (NYSE) Stock Exchanges.

Technically, both the LSE and NYSE date from the early years of the nineteenth century. Both were built on earlier institutional structures, however, and in the case of the LSE the history of the marketing of securities already dated back more than a century and had achieved considerable technical sophistication before the Stock Exchange was formally set up. According to Morgan and Thomas (1962, 11),

There are several conditions that must be fulfilled before a specialist group of dealers in stocks and shares can arise. Before there can be the makings of a regular market there must obviously be a considerable volume of securities, the ownership of which is fairly widely distributed; there must be a sufficient number of wealthy men and women who wish to hold some of their possessions in financial assets rather than in business or real property; the development of the market will also be greatly assisted if the law recognises simple procedures for the transfer of titles; if shares are of convenient and fairly small denominations, and if there is a banking system to provide a simple means of payment. All these conditions had been fulfilled in England by the end of the seventeenth century.

Throughout the late seventeenth and eighteenth centuries, a growing market for securities (principally to cover the national debt) and a smaller market for shares in joint-stock companies led to the establishment of a variety of trading arrangements. Originally, brokers and jobbers met alongside others in the Royal Exchange which, like the brokers themselves, was not highly specialized. Before they left for other venues in 1696, “dealers in stocks and shares … had a ‘walk’ near the centre of the building between the salters, the Italian merchants and the Canary merchants” (Morgan and Thomas, 1962, 27). Subsequently, the dealers met in neighboring streets and coffee houses, and in the Rotunda of the Bank of England until they were effectively evicted in the 1830s (Michie, 2001).

Although a privately-owned building called the Stock Exchange was a major site for trading from 1773, the foundation of the London Stock Exchange as an institution is usually regarded as having occurred in 1801, when access to a new building was limited to members who paid a subscription fee, creating a closed market in the midst of the otherwise open London securities market. Even then, however, the LSE did not take over the securities market, even in the capital. What it did do was to institute a system under which a self-selected group of brokers could regulate the way in which they engaged in trade. Michie therefore distinguishes between a securities market and a stock exchange by defining the latter as “A market where specialized intermediaries buy and sell securities under a common set of rules and regulations through a closed system dedicated to that purpose” (2001, 3). Thus while the LSE did not succeed in taking over the entire securities market, it did impose a framework in
which the members were able to reduce the amount of risk that they faced in comparison to outsiders who also dealt in securities.20

The establishment of the NYSE followed a similar path, although in a less developed setting. Moreover, New York’s place as the financial heart of the United States was not fully established until after the NYSE had been operating for a number of years. Until then, Philadelphia had been the dominant center. As in London, dealing in securities in New York was decentralized and uncontrolled until the last years of the eighteenth century. Brokers met in coffee houses in the vicinity of Wall Street in Lower Manhattan until 1792 when a specialized meeting place, the Tontine Coffee-House21 was built at the corner of Wall and Water Streets (across from Alexander Hamilton’s house). By then, in May 1792, a group of brokers had already agreed to regulate commissions (Sobel, 1965).

The owners of the Tontine were not specialist brokers but businessmen for whom securities were a minor interest. The establishment of the NYSE came in 1817 when the New York Stock and Exchange Board was set up following a visit by New York brokers to examine the operations of their rivals in Philadelphia (Sobel, 1965; Geisst, 2004). Like the LSE, the Stock and Exchange Board and its successor, the NYSE, had a constitution that specified who was eligible to be a member and regulated their conduct.

The main reason for establishing the two Stock Exchanges was not to protect the public but to protect the members against problems caused by other brokers. In some cases, opportunism was the target, but the rules also attempted to reduce risk. In the process, over a long period, both exchanges moved closer to what would today be considered “efficient” markets because they led to a more open distribution of knowledge among brokers.

Both markets relied on their members to be able and willing to settle their debts in full and on time. Time considerations were especially important in the case of the LSE because settlement was often weeks after transactions occurred, in contrast to the NYSE which insisted on settlement the following day to avoid any suggestion of gambling (Michie, 1987 and 1988). As commissions were a fraction of one per cent of the value of trades, the sums involved could be enormous relative to the capital of the brokers. Guarantees that members deposited with the LSE (£250 until 1832 when they were increased to £300) were tiny in comparison to the amounts that could be lost. When Thomas Manson failed in 1814, for example, he had outstanding commitments to fellow members of £134,000 while J. and L. Burnand owed £437,000 when they went bankrupt in 1835 (Michie, 2001). On the other side of the Atlantic, before the NYSE was founded, the 1792 bankruptcy of William Duer, a friend and associate of Hamilton’s, led to a decline in securities transactions that lasted for several years (Sobel, 1965). Even eighty years later, the entire operations of the NYSE collapsed with the failure of Jay Cooke and Co. in 1873 (Geisst, 2004). It is hardly surprising

---

20 Weber (2000 [1894]) distinguishes an exchange from a market on the grounds that its participants are traders acting as agents rather than members of the general public. The advantage that separates these traders from other who, in the absence of regulation, might try to deal directly in a particular security or commodity is that the professional traders have superior specialized knowledge that they are willing to share with others for a relatively modest fee.

21 “Tontine” was a reasonable name for the site of a securities market. Until the nineteenth century, shares in financial schemes such as lotteries and tontines were traded as securities, and lotteries were among the “sweeteners” that the British government used to induce people to finance the national debt (Morgan and Thomas, 1962).
that a major thrust of the regulations of both exchanges was the recruitment of sound members and the implementation of mechanisms to monitor their subsequent activities.

In Britain, the LSE formed a part of what Cain and Hopkins (2001) have called “gentlemanly capitalism”, a regime under which people of substance and sound reputation tried to restrict their business operations to others of similar background. In overseas trade, shipping, and insurance, as well in securities markets, education, family, marriage and other characteristics came to be used as sorting devices. They were not, of course, the sole indicators of probity (at least among sensible people) as shadiness in personal behavior could override background and lead to exclusion from trade with people were both cautious and of substance. Similarly, the NYSE tried to restrict its membership to people of honesty and means, although perhaps with less success than the LSE (or some other American exchanges such as Boston) were able to command (Geisst, 2004).

The membership policies of the two exchanges, although directed toward the same end, differed considerably. Until shortly before the First World War, the London Stock exchange did not attempt to restrict the number of members but concentrated instead on the characteristics of the members and imposed strict controls on their activities to reduce risk to other members. Membership was generally open to qualified members and the costs of admission were relatively cheap (around £1200 in 1904, or £440 for a stockbroker’s clerk). As a result, membership rose from 864 in 1850 to 5,567 in 1905 (Michie, 1987).22 But because the required personal wealth was small, the LSE carefully restricted the activities of its members: All partners in a firm had to be members and members were barred from outside activities including banking and the law. There were two main reasons for this. First, it was easier to monitor members who were not allowed to take part in other businesses. Equally, important was that the policy reduced the exposure of members of the exchange to misfortunes that befell non-members, a matter of great importance before limited liability became common and partners were liable not only for their own debts but for those of their partners. To reduce the chance of ripple effects arising not from opportunistic or imprudent behavior by members but from problems that might, in effect, arise anywhere in the world, the LSE simply forbade its members from becoming involved in other businesses (Michie, 1987 and 2001).

The New York Stock Exchange, by contrast, was far less open to accepting new members in later years. The original initiation fee was a very modest $25, although undesirable applicants could be, and were, blackballed. In the 1840s, this was raised to $400 in the hope of eliminating poorer applicants and from 1869 the number of seats was effectively frozen.23 From this time onwards, seats could be bought or sold and prices soon rose, reaching between $64,000 and $94,000 (£13,000-£18,800) in 1910. As a result, members of the NYSE were in general wealthier than those of the LSE. This was magnified by the fact that holders of seats in New York could engage in a wider range of business and generally formed more highly capitalized firms that were better able to withstand shocks (Geisst, 2004; Michie, 1987). Finally, the earlier availability of general limited liability in parts of the United States (Micklethwait and Wooldridge, 2003) may have reduced concerns over conducting business with outsiders.

22 Before declining to 4,855 in 1914 (Michie, 1987).

23 After merging with the Open Board in 1869, the NYSE had 1,060 members, a number that had increased by only 40 in 1914 (Sobel, 1965; Michie, 1987).
The purpose of these rules was not to establish monopolies. In fact, they probably had the opposite effect by generating membership barriers that were hard to surmount at a time when their were few in any government restrictions on trading securities in Britain or the United State. As a result, neither the London nor the New York Stock Exchange controlled trading in securities in their home markets. Both countries had regional exchanges as well as brokers who acted independently. In addition, the NYSE had two large competitors in New York itself, the Consolidated Stock Exchange and the Curb Exchange. By 1913, the NYSE had 1100 traders but the Consolidated had 1,225 and there were at least 200 members of the Curb Exchange (Michie, 1987; Sobel, 1970). Instead, the LSE and the NYSE were designed (a word used advisedly) to protect their members by regulating and monitoring their actions. In the process of reducing opportunism and risk, they also moved by stages towards the creation of what would now be known as more “efficient” markets. That is, through conscious rule-making, they were able to reduce transaction costs to manageable levels. They therefore not only illustrate the value of transaction costs (in this case the compliance costs of conforming to rules intended to mitigate the effects of opportunism and risk) in lubricating a market but also demonstrate how constructed markets can, under suitable circumstances, be alternatives to both unstructured markets and hierarchies.

4. Discussion and Conclusion

The analysis and case studies I have presented demonstrate that transaction costs are not necessarily frictions or noise and that they may actually contribute to the successful use of markets. As a concept, this is not entirely new (Macneil, 1981; Lindenberg, 2000; Nooteboom, 2002; Jacobides, 2005; Langlois, forthcoming), but I have gone further by describing how markets may be specifically designed, not only to reduce transaction costs, but to improve their economic efficiency at the same time. This is possible because for markets (in contrast to the goods or services traded in markets), what outsiders may perceive as transaction costs are frequently also production costs. As a result, it is sometimes possible for the people who control markets to reduce the transaction costs of their users in the process of reducing their production costs. Not all markets are amenable to this type of treatment because they are too decentralized and approximate Salter’s description of a “system [that] works itself” (or Adam Smith’s Invisible Hand). In some cases, however, including important financial and commodity markets, it has been possible to improve consciously on trading systems that have evolved naturally.

Although the terms regulation and self-regulation have been used in the paper, management better describes the process that I have been examining. Financial markets, such as stock and commodities exchanges, are best thought of enterprises producing services in the form of transactions. This may be done with varying degrees of efficiency, and if efficiency levels are too low, the enterprise is likely to fail. As recent attempts at diversification and merger

24 For example, in its early years, the London Stock Exchange explicitly declined to deal in foreign securities. These continued to be traded at the Royal Exchange until a Foreign Stock Exchange was founded in the early 1820s. This was then absorbed shortly afterwards into the LSE (Morgan and Thomas, 1962; Smith, 2003).

25 A major difference between the cases discussed here and by MacKenzie and Millo (2003), on the one hand, and by Jacobides (2005), on the other, is that the type of outsourcing described by Jacobides has not led in many to the creation of “manageable” markets, even though the trading undertaken is often subject to external regulation. Michie (2001) and Weber (2000 [1894]) may provide a clue to these differences when they define exchanges to involve deals among traders (agents) rather than among producers and final customers. It will be interesting to see if new communications methods change this is the case of mortgage brokers, for example, who need to maintain close contact with customers in highly diverse circumstances.
have shown, these markets are subject to many of the management techniques normally come
to mind in relation to other types of firms. In the cases discussed here, the emphasis has been
on increasing efficiency through improving knowledge flows while simultaneously reducing
exposure to opportunism and risk. This is a common program for managers in many
industries, but the need to develop coping techniques probably arose earliest in financial
markets, overseas trade, and similar areas in which the connection between communications
and opportunism and risk have long been important for survival.

Several types of lessons can be derived from the paper. The first concerns transaction cost
analysis and the relationship between transaction costs and production costs. Regardless of
whether one accepts the use of the word “trust” to describe calculative behavior, it is clear
that people have frequently consciously manipulated their environments to reduce the
probable effects arising from opportunism. In addition to markets as conventionally
understood and hierarchies, it is possible to construct markets that are as efficient as, or more
efficient than, markets that have evolved organically. This can be – and in important cases,
has been – done through either cooperation of fiat.26 This implies that economists could
profit from a deeper study of the ability of a wide range of institutional arrangements to
attenuate transaction costs. Even though there are costs involved in setting up new types of
institutions, this is not relevant as long as the resulting savings are larger. Thus further
research should extend beyond exchanges to look at arrangements between other
configurations of buyers and sellers (e.g. real estate markets and electronic auctions) as well
as along supply chains (as Cacciatori and Jacobides (2005) have recently begun to explore).

The relationship between transaction costs and classical costs of production also deserves
more thought than it has received to date. It is clear that some of the most important classes
of transaction costs have had implications for production costs that go beyond simple make-
buy decisions. For instance, the potential for hold-up is not only a possible transaction cost
but may affect the choice of a production function. When hold-up is likely, therefore, it
should be included in discussions of production costs as well as of opportunism. This
blurring between production and transaction costs extends, as here, into other parts of the
economy – in fact, into any activity in which the conduct of transactions is a major aspect of
a firm’s business. Defined in this way, transaction costs as production costs affect significant
sectors of modern economies, both for specialist traders and for firms such as McDonalds for
whom the production and the sale of a good or service are closely related.

Finally, the paper raises questions concerning the concept of “performativeness”. The term is
derived from linguistics in which “an utterance that does and says what it does
simultaneously (for example, ‘I declare this meeting open’) is termed “performative”
(Cochoy, 1998, 218, n. 1). More recently, the idea has been extended to investigate the
influence of economics as a social science, leading Callon (1998, 2) to contend that
“economics, in the broad sense of the term, performs, shapes and formats the economy rather
than observing how it functions”. If this is true, then many economic institutions may have
evolved along a path that economists suggest is somehow superior, rather than by following a
set of “natural” rules for the conduct of economic activity. It is quite possible that
performativeness represents a form of wish fulfillment for many economists, especially

---

26 In nineteenth century France, Germany, and Austria, for example, the membership and rules of stock
exchanges were established by government regulation rather than privately as in English-speaking countries
(Weber, 2000 [1894]). Whether these institutions were as well run as in Britain or the USA is a matter of
dispute (Weber preferred the English model), but they did permit the establishment of markets of sufficient
efficiency to allow widespread trade in government securities and private shares.
advisers in high places, but it is also clear that the concept can be taken too far, largely because much economic activity is, for one reason or another, not capable of being managed in this way. Nevertheless, as MacKenzie and Millo (2003) have shown for the development of derivatives trading on the Chicago Board Options Exchange since the early 1970s, performativeness can play an important role in constructing a new market in which the rules are written to elicit the outcomes predicted by a theory even though those outcomes may never have occurred before and are unlikely ever to occur in an unmanaged market environment. The development of early exchanges suggests, however, that the process is more complicated that Callon argues. These exchanges and their rules were established before the Theory of the The Market had been laid down. Instead, they had to work out their practices through trial and error, practices that were only subsequently codified into theory. These then led to the understanding of how markets should function that Black and Scholes (1973) and Merton (1973) built on in their early work on derivatives. Thus the relationship between theory and practice has been iterative and economists have learned from observation and not only developed theories deductively before broadcasting their lessons so that others might avoid mistakes. As economic conditions are continually changing, it is reasonable to expect this iterative pattern to continue.

27 In part, this is a question of the size of the installed base relative to the amount of new information that becomes available. As economic theory was rudimentary in the eighteenth century, it is not surprising that the lessons derived from the experience of the new exchanges led to larger incremental advances in theory than might be expected in the current environment in which much has already been learned and codified. Important changes in circumstances, however, might still generate important changes in basic theory, with the abandonment of some ideas that are no longer useful and their (eventual) replacement by new concepts that seem to be more valuable in explaining the altered economic environment.
REFERENCES


