Abstract

This chapter addresses the history, use and significance of the term transaction costs. Few words in the economic language have been more abused or fought over and this is shown to result from the emergence of two distinct definitions and uses. The ‘Neoclassical’ definition rests on the costs of trading across a market, while the ‘property rights’ definition centers on the costs of establishing and enforcing property rights. In articulating these two separate definitions and in demonstrating their relationship and separate uses, it is hoped that more progress can be made in the field of transaction cost economics.

JEL classification: K0, L0, L2, D0, D8

Keywords: Transaction Costs, Property Rights, Coase Theorem

1. Introduction

Transaction costs. Do another two words exist in the economic lexicon that generate as much friction? Conceptually introduced in Coase’s 1937 paper ‘The Nature of the Firm’ as simply ‘the cost of using the price mechanism’ (Coase, 1988, p. 38), the words ‘transaction costs’ have evolved to the point where some skeptics claim they include any cost that is convenient and elusive enough to avoid critical examination (Niehans, 1987, p. 678). Advocates, on the other hand, have hailed the recognition of these costs as revolutionary and as important conceptually as ‘marginalism’ and ‘substitution’ (Cheung, 1983, p. 21).

The ambiguity that surrounds the concept of transaction costs stems, in large part, from the existence of two literatures simultaneously claiming ownership over the term. The ‘property rights’ literature begins with Coase and has consistently focused on the role transaction costs play in determining the distribution of property rights, broadly defined as all laws, rules, social customs and organizations that generate incentives for behavior. This literature has called into question fundamental concepts like efficiency and the nature of production. Though based in neoclassical economics, this literature has evolved beyond the neoclassical model and has produced the new sub-fields of ‘law and economics’, the ‘new economic history’ and the ‘new institutional economics’.
Though this field, through Coase, claims the discovery and rightful title to `transaction costs', ironically the words are conspicuously absent from many of its titles. Indeed this literature is mostly responsible, though not solely, for the plethora of terms that either substitute for or refine the notion of transaction costs.

The ‘neoclassical’ literature on transaction costs begins in the early 1950s, although some might argue that it starts with Hicks (1935) or even Coase (1937). This literature defines transaction costs more narrowly, generally models them more explicitly and often analytically identical to transportation charges or taxes. The correspondence with familiar costs carries over to the types of issues examined, such as the effect of transaction costs on the volume of trade, abilities to arbitrage, the bunching of transactions, intermediation and the existence and efficiency of equilibrium - all standard neoclassical fare. Sometimes this literature examines issues of property right determination, such as the role of middlemen and the medium of exchange. In addition to the different approach and definition, the conclusions are often opposite from the property rights literature as well. This is especially true over questions of efficiency and this has increased the level of belittling rhetoric between the two camps. For example, it is common in the neoclassical literature, when reference is made to the Coase Theorem - the cornerstone of the property rights literature - to say ‘the so-called Coase Theorem’ (See Niehans, 1987 p. 678, for an example). The property rights literature is just as aggressive, claiming that the neoclassical camp often wants their cake and eat it too. For example, early criticisms over the monopoly model almost mocked the inconsistency of having a monopolist know its demand curve at zero costs, yet find it prohibitively costly to price discriminate (see Demsetz, 1969, or Barzel, 1977, for examples).

The likely cause of this dichotomous literature is twofold. First, there is the early introduction of costly transacting by Coase (1937) in the explicit context of institutional choice, at a time when the profession had little interest or ability to grapple with the issue. As Coase (1972) noted, his 1937 paper on the firm was often cited, but was little used. Second, there is Coase’s failure in 1937 to define transaction costs with any precision, using instead the phrase ‘the costs of the price mechanism’. At the same time, though Coase uses examples that suggest more than just the market is involved in transaction costs, he ultimately leaves the issue open for interpretation. As such, the property right literature did not truly begin until 1960, with Coase’s publication of ‘The Problem of Social Cost’. This latter article provided the necessary elaboration of Coase’s 1937 publication in order to tie many existing ideas together and to provide a property rights research agenda (see Barzel and Kochin, 1992, or Medema, forthcoming, for elaborations on this point). In the intervening years, economists did what they could with the term transaction costs and the neoclassical approach was born.

The purpose of this chapter is to provide a broad picture of transaction costs: its history, definition, foundation, use, measurement and implications.
As such, it is often necessary to sacrifice detail and the reader is directed to explore the references for further treatment. A theme throughout the chapter is the dichotomous use of the term ‘transaction costs’ in the two streams of literature already mentioned. It is ironic that a disagreement over ownership should engulf a term so closely related to property rights. Unfortunately, as with all cases of disputed ownership, useful output is lower for lack of definition.


In the beginning Coase created transaction costs. His critics might continue: ‘And the term was formless and void and darkness was over the surface of the term’. For the believers in the property right approach, however, Coase (1937) is seminal. As an advanced undergraduate perplexed by economics’ ability to conceptually organize the economy around prices, Coase was troubled that there was no room for any form of direct cooperation or direction. In his words ‘we had a factor of production, management, whose function was to coordinate. Why was it needed if the pricing system provided all the coordination necessary?’ (1992, p. 715). His solution was to recognize that there are ‘costs of using the price mechanism’. When prices allocate resources at a cost, then they compete with other allocating mechanisms like firms and governments. Coase argued that, at times, firms and direct management supersede the market, while at other times market prices are used in directing goods and services. Readers interested in the genesis and a detailed account of the history of Coase’s first great work are directed to Williamson and Winter (1991).

In this simple argument a charitable reading finds some basic elements that distinguish the property rights literature. First, all methods of allocating resources have costs and benefits and no single mechanism works for free and dominates all others - in modern language, all allocation mechanisms are ‘second best’. Second, it is argued that ‘rules’, ‘organizational forms’ and ‘methods of payments’ are subject to economic analysis. Although it has been argued that Frank Knight (1921) indirectly made a similar case (see McManus, 1975; Barzel, 1987), Coase explicitly addressed this issue. And finally, Coase implicitly argues that positive transaction costs were both necessary and sufficient for an explanation of the firm.

Coase provides examples of what he meant by the costs of the price mechanism: discovering what the prices are, negotiating and closing a contract; and he hints at problems of enforcement, but he stops short of any definition. In fact, throughout all of his writings, Coase never goes beyond providing examples of transaction costs. Barzel and Kochin (1992, p. 25) have noted that ‘the discussion of transaction costs in that [1937] paper is brief and cryptic’ and even the most sympathetic reader would have to agree. Though the words
‘transaction costs’ are never used in his first work, Coase is still correct when, in his Nobel address, he states that: ‘What I think will be considered in the future to have been the important contribution of this article is the explicit introduction of transaction costs into economic analysis’. (1992, p. 716).

It remains a strange fact of economic history that after the publication of ‘The Nature of the Firm’, neither Coase, nor any other writer in the profession, picked up the joint theme of transaction costs and property rights. Finally, in ‘The Federal Communications Commission’, Coase (1959) returns to the theme of the influence of transaction costs on property rights and this article provides the motivation for ‘The Problem of Social Cost’ (see Kitch (ed.), 1983, or Stigler, 1988, for discussions of how Coase came to write his most famous paper). Ironically, even Coase did not appreciate his accomplishment at the time of writing:

I should add that in writing this article I had no such general aim in mind. I thought that I was exposing the weaknesses of Pigou’s analysis of the divergence between private and social products, an analysis generally accepted by economists and that was all. It was only later and in part as a result of conversation with Steven Cheung in the 1960’s that I came to see the general significance for economic theory of what I had written . . . (1992, p. 717)


Cheung has made many contributions to the property rights literature on transaction costs, but perhaps his most significant is generalizing Coase’s original argument. The importance stems from the fact that Coase never defined transaction costs and has often used examples that suggest transaction costs arise only in market exchanges. Cheung, in analyzing share tenancy and providing the first contractual example of the Coase theorem, explicitly argues that contract choice depends on the transaction costs of the different contracts. These transaction costs are clearly internal and not just market costs. Cheung’s work inspired Stiglitz (1974) and begins the principal agent literature, but it also establishes the precedent of thinking of transaction costs across markets and internal to the firm - a theme that is strongly articulated in Williamson
This connection between transaction costs and property rights is summarized in ‘the Coase Theorem’, which is defined as follows:

**Coase Theorem:** In the absence of transaction costs, the allocation of resources is independent of the distribution of property rights.

There are many attacks and defenses of the Coase Theorem, none of which are dealt with here (see Shapiro, 1974, for an example of an attack, Allen, 1997, for a defense and Zerbe, 1980, for a survey). The point is that for all property right approaches to transaction costs, the two concepts of property rights and transaction costs are fundamentally interlinked. In fact, it will be shown that they are two sides of the same coin and that this linkage distinguishes the property right approach from the neoclassical approach to the study of transaction costs.

**Property Rights and Transaction Costs**

The delineation of ownership is as old as human written records. The Mosaic laws as described in the Ten Commandments or the laws on takings in Exodus 22:1-15, as well as the host of other Levitical laws throughout the first five books of the Old Testament, are all attempts to legally define ownership. From the Hammurabi code to the English common law the notion of legal ownership, or legal rights, to property is well defined. In the words of Blackstone: ‘The third absolute right; inherent in every Englishman, is that of property: which consists in the free use, enjoyment and disposal of all his acquisitions, without any control or diminution, save only by the laws of the land’ (1803, p. 138).

Though it is difficult to identify where one idea begins, the modern attempt to go beyond a legal delineation of rights and begin talking about ‘economic rights’ seems to start with Alchian. Alchian’s early work on tenure (1958) and the pursuit of individual utility within the context of regulated firms (Alchian and Kessel, 1962) hinge on the property right structures of the institutions in question. For example, managers and administrators of non-profit firms and universities, he argues, face a lower relative cost of private consumption on the job than their counterparts in the private sector. Because these firms are constrained in their ability to show profit, they are able to survive with higher costs. Alchian’s insight was that the set of rules (the distribution of property rights) determined the level of output of the firm because they determined the incentives of each individual. This theme is manifest throughout Alchian’s work and culminates in his famous article with Demsetz (Alchian and Demsetz, 1972). But perhaps Alchian’s most significant contribution, articulated most clearly in Alchian (1965, 1979), is his emphasis on economic rather than legal rights. For Alchian, property rights are ‘the rights of individuals to the use of resources’ (1965, p. 817) not just under the law, but in reality. He makes clear that these rights are not solely dependent on the
existence of the state, but that they depend on custom, reciprocity and voluntary restraints. This notion is now commonplace in the modern property rights literature and is explicitly found in Ellickson (1991) and Landa (1994). Although economic property rights are enhanced by the law, they are ultimately use rights and the greater extent one can exercise these uses and bear the consequences the greater are the property rights, regardless of the law. Property rights are therefore defined as:

**Property Rights**: the ability to freely exercise a choice over a good or service.

The property rights literature argues there is a monotonic relationship between property rights and wealth. Given that trade is the transfer of property rights, there can be no trade (and hence no gains from trade) in the absence of property rights. Also, when property rights are perfectly defined, the Coase theorem states that the gains from trade are maximized. Assuming there is a continuum between these two extremes, as property rights become better defined, the gains from trade increase (see Anderson and Lueck, 1992 for an empirical example). Other things equal, individuals prefer better defined property rights to poorer defined ones because they prefer more wealth to less.

Increasing the ability to make choices of one individual can reduce the ability to make choices for others. Generally speaking individuals increase their property rights in three ways. First, the individual may steal the good in question. Second, the individual may privatize a good that was previously in the public domain. Finally, an individual may cooperate with other individuals with an agreement to divide the new wealth in some fashion.

When property rights are perfect, by definition no theft can take place and as a result, no effort is made to protect the rights (a point made in Cheung, 1974 and Barzel, 1985). However, when property rights are incomplete, individuals attempt to increase their ownership in an effort to increase their wealth. This attempt to capture property rights may be dissipating (as in the case of theft), or may be wealth generating (as in the case of assets brought out of the public domain). When there is an opportunity for theft, there is also an opportunity for protection. Hence, when property rights are incomplete, individuals are always in the process of maintaining their existing property rights and attempting to establish new ones. This leads to the property right definition of transaction costs.

**Transaction Costs, #1**: the costs establishing and maintaining property rights.

This definition is first articulated in Allen (1991). Writers in the property rights literature have seldom defined transaction costs, relying mostly on
examples of inspection, enforcing, policing and measurement which all hint at
the protection of property rights and implicitly recognize the threat of
appropriation or theft. For similar, but informal, definitions, see Cheung (1969,
p. 16), McManus (1975, p. 336), Jensen and Meckling (1976, p. 308), Barzel

Transaction costs include any direct costs, as well as any concomitant
inefficiencies in production or misallocation that resulted from them. For
example, consider the Klein and Leffler (1981) example of a firm investing in
a sunk asset as a guarantee of product quality. The firm does this to protect
the wealth of its customer and as such it is clearly an attempt to maintain property
rights. The transaction costs would include the cost of the investment and any
increases in costs of production that it may have caused.

The property rights definition of transaction costs respects no boundaries
between firms, markets, households, or any other theoretical constructs. When
property rights are protected and maintained in any context, transaction costs
exist. By explicitly recognizing this relationship it is clear that statements like
‘if we assume zero transaction costs and complete property rights’ are
redundant. To say that a situation has zero transaction costs is to say that
property rights are complete, according to this definition. Cheung (1992, p. 54)
gives this, stating: ‘the dual specifications of clearly delimited rights and
zero transaction costs are redundant. If transaction costs are truly zero, the
delineation of rights can be ignored’.

When it is costless to establish and maintain rights they are done so
perfectly. If transaction costs are prohibitively high then property rights will
neither be established nor maintained and property rights will be zero. The
reverse, however, is not necessarily true. If property rights are complete in some
situation, there are two possibilities, either transaction costs are zero, or costs
may have been incurred to guarantee the property rights simply because the
benefits of doing so exceed the costs - in which case transaction costs are
positive. Further, when property rights are zero, transaction costs could also be
zero. For example, if a property right could never be established, despite the
resources devoted towards such a goal, no one would bother making any
expenditures towards establishing property rights and the good would remain
unowned. For example, there are no property rights over the planet Venus and
no efforts have been made to establish any.

Transaction Cost Economics with the Property Rights Approach
An excellent survey of the property right literature is found in Eggertsson
(1990a), while an excellent textbook treatment of this approach is found in
Milgrom and Roberts (1992). Essentially the property rights literature is
characterized by several features related to the above definition. First, the
central question is always ‘what explains the distribution of property rights?’,
where the ‘distribution of property rights’ has a broad meaning and includes all sets of rules, governance structures and organizations. Hence, families, firms, governments, non-profit institutions, contracts, are all viewed as sets of property rights. Lawyers forming a partnership to split the residuals, a farmer renting land from a landowner, or a judge deciding on a case, are all examples of different allocations of property rights. Every distribution of property rights has with it a set of production costs and a set of transaction costs. The distribution of property rights that maximizes the gains from trade net of all costs is the optimal distribution. This, in fact, is the grand hypothesis of transaction cost economics under the property rights approach. An account of transaction cost methodology is beyond the scope of this paper, but see Williamson (1979, 1985) for detailed accounts.

A second characterization is the reluctance to infer any policy implications from the analysis and to stress explanation. As stated earlier, this goes back to Coase’s original idea that no single allocation mechanism dominates. Notions of ‘market failure’ lose meaning when there is no reason for prices to allocate everything. One might as well refer to ‘government failure’ or ‘firm failure’ in cases where prices do allocate.

This transaction cost approach dominates what is now called the ‘New Institutional Economics’, so named because it provides a theoretical framework and emphasis of testability to the institutional traditions of Veblen and Commons. Oliver Williamson is considered the founder of this literature, both in terms of vocabulary and content and he is one of the strongest proponents of applying the notion of transaction costs ubiquitously. His notion of a ‘governance structure’ as a distribution of property rights providing appropriate incentives to govern a relationship, is intended to apply within and outside firms. Williamson (1971) is the first to note the role sunk costs can play in causing contracting problems and incentives to vertically integrate. This idea is popularized in Klein, Crawford and Alchian (1978) and in Klein and Leffler (1981). The role of asset specificity and idiosyncratic capital is so attached to the name of Williamson that for many, transaction costs means little else. Although Williamson’s understanding of the relationship between transaction costs and property rights is consistent with what is presented here, he also distinguishes between the ‘property rights approach’ and the ‘transaction cost approach’ to organizational problems. For Williamson, a property rights approach deals with grand private environmental rules, while the transaction cost approach deals with private incomplete contracts (see Williamson, 1990 for a discussion).
3. The Tale of Two Histories, Part B: The Neoclassical Approach

Although, Coase (1937) provides mostly market exchange examples and could be argued as the founder of the neoclassical approach to transaction costs, it could be better argued that this approach begins with Hicks’ (1935) publication ‘A Suggestion of Simplifying the Theory of Money’, which predates Coase by two years. In his paper, Hicks begins what is known as a transaction demand for money, although he never calls it as such. For him, there are frictions in the economy and these apply to buying and selling capital assets yielding positive returns. When the returns were small, at the margin, relative to the costs of trading, individuals rationally hold cash balances yielding no return. In his words:

The most obvious sort of friction and undoubtedly one of the most important, is the cost of transferring assets from one form to another. This is of exactly the same character as the cost of transfer which acts as a certain impediment to change in all parts of the economic system; it doubtless comprises subjective elements as well as elements directly priced. Thus a person is deterred from investing money for short periods, partly because of brokerage charges and stamp duties, partly because it is not worth the bother. (1935, p. 6)

Since money is used to facilitate exchange and since an exchange that needs ‘facilitating’ must be subject to transaction costs, it is not surprising that those concerned with money dealt with these costs. Indeed, Baumol (1952) and Tobin (1956) elaborate on the transaction demand for money and again treat transaction costs as the costs of trading. The first explicit statement of transaction costs as the cost of trading comes from Demsetz (1964) where he states that ‘Transaction cost may be defined as the cost of exchanging ownership titles’ (1988, p. 64). Although this type of definition refers to property rights, transaction costs only arise when an exchange of property rights takes place. This leads to the neoclassical definition of transaction costs:

**Transaction Costs #2:** the costs resulting from the transfer of property rights.

Timmermann (1994), Shaffer (1989), Stavins (1995), Wagner and Schulman (1994), Wilcox (1993) and Young (1989). A typical definition of transaction costs found in these papers would be as follows:

In general, transaction costs are ubiquitous in market economies and can arise from the transfer of any property right because parties to exchanges must find one another, communicate and exchange information. There may be a necessity to inspect and measure goods to be transferred, draw up contracts, consult with lawyers or other experts and transfer title. Depending upon who provides these services, transaction costs can take one of two forms, inputs or resources - including time - by a buyer and/or a seller or a margin between the buying and selling price of a commodity in a given market. (Stavins 1995, p. 134)

In the neoclassical approach, enforcement-type costs within firms are not transaction costs. Transaction costs consist of those costs that occur between firms or individuals from the process of market exchange. Hence, an economy made up of one giant firm, or a state run economy, would be a zero transaction cost economy by this definition. Because these transaction costs are just the cost of exchange, they are modeled in a more recognizable fashion, often in the form of a ‘transaction function’ (see Constantinides, 1979 for an example). These functions are similar to other neoclassical production functions and are usually assumed to depend on labor inputs. These functions may have increasing, constant, or decreasing returns to scale. Further, the transaction cost functions may have fixed or variable components. Although the analogy is not complete, in many ways transaction costs play a role very similar to transportation costs and taxes and, according to Niehans: ‘transaction costs are analytically analogous to transportation costs’.

Being analytically similar means that many of the impacts of transaction costs are similar as well. Consider, for example, the impact of transaction costs on the volume of trade. If transaction costs increase with the quantity traded, this has the impact of increasing the relative price of the commodity being purchased. Since this holds for goods, in effect the budget constraint becomes kinked at the endowment point and, as a result, individual demands become less responsive to price changes and the volume of trade falls. These are often called ‘proportional transaction costs’ in the literature and their effect on multiperiod investment and consumption has also been examined. (See Bensaid et al., 1992; Boyle and Vorst, 1992; Constantinides, 1976; Davis and Norman, 1990; Eppen and Fama, 1969; Kamin, 1975; Leland, 1985; and Magill and Constantinides, 1976). Other similar results follow as well. Like per unit taxes, frictional per unit transaction costs drive a wedge between buying and selling prices, although neoclassical transaction costs are not necessary to explain price spreads. Glosten and Milgrom (1985), based on Copeland and Galai (1983),
provide an adverse selection explanation for bid-ask spreads that assumes traders have zero friction costs.

Fixed transaction costs tend to bunch transactions together and provide an explanation for the demand for money (see Edirisinghe, Naik and Uppal, 1993, for an example). Differences in transaction costs across individuals lead to some specializing in the transaction function. Hence brokers and agents are those individuals with low transaction costs. Alchian and Allen (1964) were probably the first to note this (see also Niehans, 1969). Differences in the transaction costs across commodities provide an explanation for why some commodities are used as currencies of exchange (Niehans, 1969 and Alchian, 1977). In these last two cases, the question examined is close to the institutional type of question addressed by the property rights school. Neoclassical transaction costs have also been used to analyze the equity premium. The real average returns on US Treasury Bills is less than 1 percent, while for stocks it is closer to 7 percent. This difference is too large to explain with reasonable Arrow-Debreu models. Mehra and Prescott (1985) began a literature explaining this premium based on neoclassical trading costs. (See Aiyagari and Gertler, 1991, for an example and a survey of the literature.) Finally, all discussions of the existence of equilibrium with transaction costs utilize a neoclassical definition (See Bergstrom, 1976; Foley, 1970; Hahn, 1971; Hart and Kuhn, 1975; Heller and Starr, 1976; Kurz, 1974b; McKenzie, 1981; Radner, 1972; and Repullo, 1988).

**Definitional Squabbles**

For the most part, these two streams of literatures - the property rights approach and the neoclassical approach - flow independently. Those writing in the area of property rights follow the line of reasoning laid by Coase, Cheung and Williamson and use the broad notion of transaction costs. Those interested in the neoclassical issues of volume of trade and equilibrium generally stick to an Arrow-Debreu based general equilibrium model and use the narrow definition of straight exchange costs.

The major exception is Harold Demsetz. Demsetz was an early contributor to the theory of property rights and the role of enforcement costs in determining the distribution of property rights (See Demsetz, 1964, 1967 and 1972). Ironically though, he was also the first to articulate the neoclassical definition of transaction costs (Demsetz, 1968). For Demsetz, transaction costs remain ‘the costs of coordinating resources through market arrangements’ (1995, p. 4) and among property right economists he remains a staunch, though perhaps lonely, proponent of this view. Demsetz (1964) is the first to deal with the breadth of definition used for transaction costs. In Demsetz (1988) he acknowledges that this is mostly a question of semantics, since his collection of costs all fit under the rubric of ‘governance’ costs or the property rights definition. According to Demsetz, the clear meaning of transaction costs is the
cost of transacting. To apply the term more broadly threatens to make the term
tautological and useless. This view is summarized by Schlag, 'an overly
expansive view of transaction costs threatens to make the Coase theorem
tautological. On the other hand, an overly restrictive view of transaction costs
can effectively invalidate the theorem' (1989, p. 1675).

Demsetz (1988, pp. 144-150) argues that a broad definition of transaction
costs hinders any understanding of firms and markets. For example, Demsetz
(1995) argues that the definition of a firm and its internal organization are two
separate issues that have been confused since Coase. Demsetz defines the firm
as a production unit, created to exploit gains from specialization. Since markets
only transfer titles they complement firms and the Coasean notion of firms and
markets substituting for one another does not arise. This is exactly the opposite
way the property rights literature would define a firm (see Barzel, 1989, as an
example). Coase has always put emphasis on the formal relations within a firm
(for example, employer vs. employee) as a possible means of reducing
transaction costs. Alchian and Demsetz (1972), on the other hand, downplay
the role of authority within the firm. (See Medema, 1994, for a discussion of
Coase vs. Alchian and Demsetz.)

Demsetz, de facto, takes a property rights approach to the internal
organization of the firm, however. Demsetz (1995) discusses several transaction
costs (definition #1) arguments for the firm without using the term, including:
shirking, Knightean uncertainty, reduction of coordination costs and the agency
problems from opportunism. Hence, in the end there is very little to quibble
over and the definition to be used depends on the problem being addressed.
Clearly, all of the costs mentioned by Demsetz fall under the umbrella of the
property right definition of transaction costs, where a broad transaction cost
definition is necessary in order to make clear that the Coase theorem does not
apply.

4. The Distribution of Property Rights vs. The Volume of Trade

The economics profession is littered with various assertions and theorems
stating that distributions of property rights do not matter. The Coase theorem
is the most famous of these, but there are many others. For example, the
Modigliani/Miller theorem (1958) is almost identical to the Coase theorem.
This proposition states that if capital markets are perfect and firms and
investors face the same rate of interest, then investors can unravel any corporate
structure chosen by the firm. This means that the ratio of debt to equity
financing, as well as the form of debt and equity within the firm, is irrelevant
to the firms value. A similar result is found in the Ricardian Equivalence
Theorem (1951). This theorem states that with perfect capital markets, the
government’s choice over taxation and debt is irrelevant to the level of
household wealth, because taxpayers are able to unravel any financing decisions of governments.

In addition to these, there are a host of equivalence results regarding taxes, the most famous being that it is irrelevant whether the consumer or the producer is taxed, the result is the same in terms of both resource allocation and incidence of tax paid. Furthermore, both ad valorem and per unit taxes are equivalent in terms of resource allocation. Finally, in trade policy and again in terms of resource allocation, it is well known that tariffs and quotas can have identical effects.

All of these results are special cases of the Coase theorem because all taxes, debt obligations, equity shares and other policy instruments are delineations of property rights. A firm deciding on the optimal amount of debt versus equity is essentially assigning property rights over the stream of expected profits, including priority in case of unexpected shortfalls. A government deciding on a choice between taxation and debt is simply transferring property rights over time. In all of these cases, only a different distribution of property rights exists and given the Coase theorem, this does not alter the allocation of resources - when, as Coase stated, transaction costs are zero.

When transaction costs are not zero, these equivalence results do not occur. For example, Barzel (1976) shows how the tax equivalent result is altered when transaction costs are positive. When goods are complex bundles of commodities they become difficult to define under tax legislation and some attributes are possibly ignored. Under these conditions, taxes have the effect of altering the relative price of the taxed and untaxed attributes and therefore alter the mix or quality of the item that is produced. Lump sum taxes tend to increase the quality of good, while per unit taxes tend to lower quality. The result is that different forms of taxation can have vastly different effects on resource allocation. Furthermore, differences in the ability to avoid taxation implies that tax revenue is not neutral with respect to the location of the tax.

All these examples explain why the property rights approach requires a broad definition of transaction costs. Given the Coase theorem and all of its different manifestations, distributions of property rights are irrelevant. If we did live in a world of zero transaction costs (definition # 1), then firms truly would toss coins to decide debt levels, if indeed there were any firms, which would also be decided by a coin toss. And so on for governments and all other institutions. The importance of the Coase theorem then is that it points to transaction costs as the necessary factor in any explanation of the distribution of property rights. The definition of transaction costs, therefore, must be those costs that cause the Coase theorem to not apply. This also seems to be the reason why the neoclassical approach never analyzes questions of economic organization outside of the choice of medium of exchange. They have selected a definition of transaction costs that is too limited for this purpose. Many in the
neoclassical literature have balked at this line of reasoning, suggesting that it is tautological. The difference of opinion stems from the different objectives each approach is interested in.

5. The Causes of Transaction Costs

Regardless of which stream of literature is examined, the underlying theme for transaction costs is the notion of ignorance. Hence, even though its treatment is different and the definition is narrower, the neoclassical approach still uses examples of transaction costs that are similar to the property rights approach. Niehans states that parties must find each other, they have to communicate and to exchange information . . . goods must be described, inspected, weighed and measured. Contracts are drawn up, lawyers may be consulted, title is transferred and records have to be kept. In some cases, compliance needs to be enforced through legal action and breach of contract may lead to litigation. (1987, p. 676)

Negotiation, fraud, communication and contract stipulation all come about because knowledge is incomplete and not common. Though its importance is recognized by everyone, the role of information leads to a great deal of confusion in the discussion of transaction costs. Information costs are a prerequisite to transaction costs and are a necessary condition for their existence. Information costs, however, are not always transaction costs. Steven Cheung once remarked that transaction costs are costs that do not exist in a Robinson Crusoe world (a definition consistent with definition #1). Clearly Crusoe faced many information problems, but until Friday showed up, he had no transaction cost problems.

Barzel has been a strong proponent of the distinction between information and transaction costs. Barzel (1977) states that ‘transaction costs include those (costs) required to formulate and to police contracts’ (p. 292), but goes on to point out that it is possible to have information problems resulting in speculation, sorting and signalling, which may appear to yield decreases in social value, but that these reductions are impossible when transaction costs are zero. With zero transaction costs, contracting is a perfect substitute for information because contracts can always be made over all contingencies. Barzel (1982, 1985) stresses that information costs are at the heart of transaction costs because they lead to measurement. Barzel (1977) notes that when the distinction between information costs and transaction costs is made, several other points follow rather obviously: costless information implies perfect property rights; individual honesty does not necessarily eliminate transaction costs; costly information means transaction costs can explain
self-imposed constraints; and total costs, not just transaction costs or information costs, are required to be minimized.

It is not always appreciated that information costs are not sufficient for transaction costs. The mere presence of information costs lead to risky events which can be eliminated through contingent claim contracts. In addition to costly information a factor is required to eliminate the ability to write complete contingent claim contracts. There are several examples of what this factor might be. Knight was the first to suggest this with his distinction between risk and uncertainty; uncertainty arising in situations where moral hazard prevented individuals from assigning accurate probabilities to events and thereby eliminating the ability to contract over the risk. Barzel (1989) and Allen (1991) have stressed the idea that goods are complicated bundles of attributes that both are variable in nature and alterable by individuals. The inability to separate the contributions to quality by nature and man allows for cheating to take place in equilibrium. Other attempts to add to information costs include the notion of asymmetric information and opportunism (see Ackerlof, 1970, and Williamson, 1975, 1985).


As may be expected, the two literatures have different methods by which transaction costs are modelled. In both cases, transaction costs considerably complicate the neoclassical model and the level of mathematical sophistication is quite demanding. The major point made here, however, is that neoclassical modeling is a direct extension of the Arrow-Debreu model, while property rights modeling involves some fundamental differences.

In an Arrow-Debreu world with complete contingent markets, trades only take place once. An early application of transaction costs in neoclassical models explained why markets had a ‘sequence’ over time - the general idea being that at any given time, a specific market may be too expensive to trade in and thus trade is postponed until some future date.

There are two general types of approaches in modeling neoclassical transaction costs. The first, used by Foley (1970), Hahn (1971, 1973) and Starrett (1973), involves a central transactor who takes buy and sell orders from each household and carries them out. In order to pay for his services, the ‘broker’ charges a margin between the buying and selling price for his efforts. The second approach requires households and firms to directly use resources in the purchase and sale of goods. Here the firms and households use some type of ‘transfer technology’. (For early treatments, see Kurz, 1974a; Niehans, 1971; or Ulph and Ulph, 1977. See Repullo, 1988, for a later treatment using this
Although the specific technologies are generally simple, they are usually sufficient to complicate the analysis greatly.

Three general approaches are taken to model the transaction technology. The first simply assumes that some general transaction function \( T(x) \) exists (see Brennan and Copeland, 1988, for example). This function is often assumed to depend on the volume of trade, cash flows, number of traders and other such variables that reflect the 'size' of the transaction. The second assumes that transaction costs are fixed (see Leland, 1974; Mukherjee and Zabel, 1974; Brennan, 1975; Goldsmith, 1976; Levy, 1978; or Mayshar, 1979, 1981; for examples). Finally, proportional (or 'iceberg') transaction costs \( k(x) \) are assumed, where \( k \) is a constant fraction and \( x \) again is a measure of the size of transaction (see Gennette and Jung, 1994, or Constantinides, 1986, for examples). All of these technologies make their way into standard objective functions for firms and households. Though the subsequent analysis is usually complicated, the results are most often exactly analogous to the effects of transportation charges. Typically, these analyses show that the presence of transaction costs reduces the frequency and volume of trade.


The property rights approach to modeling is a vast, diverse and technically complex literature, well beyond the scope of this survey to treat it in any detail (see Holmstrom and Milgrom, 1994, or Hart and Moore, 1990). Unlike the neoclassical literature, where transaction costs enter and yield results which are somewhat predictable, modeling the distribution of property rights is fundamentally different. Rather than entering through a transaction technology, transaction costs arise through changes in incentives and manifest in changes in values in different property right distributions, with often surprising results. For example, Coase (1960) is perhaps the first surprising result, despite the lack of formalization. Cheung (1969) is another and perhaps the first case of a formal treatment of transaction costs from a property rights approach. Here two examples of property rights modeling are provided to highlight some differences.

The simple example of insurance, first discussed in Rothschild and Stiglitz (1976), demonstrates some differences. Consider a world where there are two types of behavior: careful and uncareful and all else equal, individuals prefer being uncareful. Furthermore, there is the chance that a fire may occur and the probability of this event depends on the behavior of the individual. If insurance companies can fully observe behavior they offer a full insurance contract and everyone takes it - no one has an incentive to be careless.
Thus far, we have a standard neoclassical problem and the introduction of risky events has changed little. Recall that the marginal rate of substitution of the individual for wealth in both states of the world is:

\[
\frac{\pi_a}{1 - \pi_a} \frac{U'(W_a)}{U'(W_b)}
\]

where \( p_a \) is the probability of fire in state ‘a’, \( U \) the individual’s affine utility function and \( W \) the wealth level in the two states. In the case of pure uncertainty the probabilities are determined by nature.

However, if the behavior of the individual is not observable, the probabilities are alterable by the individual and a transaction cost problem arises. As has been noted above the transaction cost problem requires: (1) the presence of uncertainty (here the probability of a fire) and (2) the ability of the individual to change his behavior without costless detection. Since the firm cannot observe behavior, this implies individuals all become careless, which alters the marginal rate of substitution! The introduction of costly information leads to preferences no longer being fixed and exogenous and this is an example of a fundamental difference between the two types of models. (Arnott and Stiglitz, 1988) explore the implications of shifting marginal rates of substitution.)

The solution to this particular problem has the insurance company offering an incomplete contract (an insurance contract with a deductible), which points to a second difference. Namely the possible non-existence of explicit transaction costs in equilibrium. The insurance company, by offering an incentive compatible contract, does not engage in any form of direct monitoring. Such monitoring is not necessary and many property right models have no actual resources used to establish and maintain property rights in equilibrium. In this case, the transaction costs are simply the lost gains from trade that result from the incomplete contract.

As a second example, consider a variation on the principal-agent model first introduced by Stiglitz (1974). In this model the effort of a risk-averse agent is unobservable and so a contract is reached that trades off incentives for risk avoidance. For example, consider the case of cropshare contracts, where a risk-averse farmer contracts with a risk-neutral landowner (Allen and Lueck, 1995). For a plot of land, output is \( q = (e + ?) \), where \( e \) is the unobservable labor effort and \( ? \) is a random variable with mean 0 and variance \( s^2 \). Furthermore, assume that the farmer’s income is \( Y = aq + \beta \) and his utility is \( U = E(Y) + \frac{(r/2)\text{Var}(Y)}{2} \), where \( r \) is a measure of risk aversion, \( a \) is the share of output and \( \beta \) is a fixed side payment. Finally, assume that the cost of effort to the farmer is \( c_0 + c_1e + (c_2/2)e^2 \).
For a given output share the effort which maximizes farmer utility is:

$$\hat{e}(\alpha) = \frac{\alpha - c_1}{c_2},$$

which represents the behavior of the farmer and becomes a constraint to the landowner designing the optimal contract. This incentive compatibility constraint represents another example of a fundamental distinction in modeling property right distributions - namely, the constraints often involve optimization problems. The next stage in this particular problem involves the landowner maximizing his expected income $E((1 - \alpha)(\hat{e} + \beta) + \beta)$ subject to the incentive constraint and a participation constraint.

Although the principal-agent model has been extended and broadly applied, (see Dewatripont, 1989; Freixas, Guesnerie and Tirole, 1985; Holmstrom, 1979, 1982; or Shavell, 1979, for examples), it has recently fallen out of favor for models where all parties are risk neutral (see Eswaran and Kotwal, 1985; Grossman and Hart, 1986; Leffler and Rucker, 1991; and Allen and Lueck, 1992a, for examples). Holmstrom and Milgrom (1991) develop an explicit principal-agent model where risk aversion is not required. The great advantage of risk neutrality is that it allows for several margins over which transaction cost behavior can take place. However, though there remains no single way to model transaction costs in the property rights approach, the bottom line remains that it does involve some fundamental differences from putting a ‘T’ in a cost function.

8. Direct Empirical Work

and Klein, 1995, for a more complete listing of references to the empirical literature).

Studies in the neoclassical approach are also numerous and mostly focus on asset arbitrage, the volume of trade, risk adjusted returns and the bundling of transactions (see Demsetz, 1968; Fisher, 1994; Frenkel and Levich, 1975; Litzenberger and Rolfo, 1984; Malkiel, 1966; Pesaran and Timmermann, 1994; Phillips and Smith, 1980; Protopapadakis and Stoll, 1983; Schultz, 1983; Smiley, 1976). It should be stressed that the empirical transaction cost literature seriously tests hypotheses and therefore by its existence refutes the assertion that transaction cost economics is tautological. However, most of property right and neoclassical empirical studies are of the comparative static variety and attempt to test transaction cost hypotheses using various proxies for asset specificity, uncertainty, measurement costs, friction and other transaction cost variables in reduced form equations. There are only two studies that have attempted to measure the level of transaction costs.

The first and perhaps most ambitious of these is Wallis and North (1986), who attempt to measure the entire transaction sector of the economy over 100 years. Understandably, the first problem they face is how to define transaction costs. Their property rights background leads them to define transaction costs as ‘the resource costs of maintaining and operating the institutional framework associated with capturing the gains from trade’. In the end, however, they simply separate resources devoted to transacting as their measure and in doing so ironically come closer to a neoclassical definition. Although they acknowledge the conceptual problems this definition has with respect to firms, they settle for the following compromise:

We divide occupations into those that provide primarily transaction services to the firm and, by elimination, those that provide primarily transformation services. The wages of employees in these ‘transaction occupations’ constitute our measure of the transaction sector within firms. (1986, p. 100)

This compromise would require all protective services (police, courts and so on) included in the non-transaction sector of the economy, which makes Wallis and North so uncomfortable, they switch its classification (pp. 102-103). The analysis of Wallis and North concludes that the transaction sector accounts for a significant part of the economy and that this has grown from 25 percent to 40 percent over the years 1870 to 1970.

Davis (1986), however, has pointed out that this estimate is not robust for even small changes in the line that separates ‘transactions’ from ‘production’. In the end, the problem of definition seems overwhelming. Is a farmer a manager/marketing agent, or a grain-growing field hand? All jobs have elements of production and transaction in them and it seems an impossible task
to separate them. This perhaps best explains why Wallis and North were both the first and the last to tackle transaction costs on such a grand scale.

A more sophisticated treatment of measuring the costs of organization is found in Masten, Meehan and Snyder (1991). They note that much of the empirical literature proxies only ‘the hazards of market exchange’ and ignores the internal costs of governance. Reduced form estimates are unable to distinguish between internal and external transaction costs.

Furthermore, attempts to directly measure transaction costs are subject to the problems faced by Wallis and North. Finally, Masten, Meehan and Snyder recognize the selection bias that occurs since the efficient organization structure is chosen and the other choices are not observed. Their solution is to utilize switching regression techniques and to adopt censored regression models used in labor economics. From this technique they obtain actual dollar estimates of organization costs and therefore can estimate the magnitude of individual coefficients and not just their relative impact. Masten, Meehan and Snyder apply this methodology to naval shipyard contracts and find that overall organization costs amount to 14 percent of total costs. They estimate that if an incorrect contractual agreements is chosen that this would lead to increases in organizational costs of up to 70 percent.

9. Conclusion

The essential element of transaction costs, that property rights must be protected, is found in most fields of economics and throughout the discipline’s history. Adam Smith, in discussing foreign trade, endowments, corporate ownership structure and non-profit organizations repeatedly exploits concepts of costly information and the ability of individuals to exploit others’ ignorance to their own advantage (see West, 1990, for an account of Smith’s anticipation of modern economic ideas like principal-agent relations). In macroeconomics the notion of costly information lead to the rational expectations revolution and subsequent real business cycle models based on search and the disincentives found in unemployment insurance programs. Public choice models are founded on the premise that individuals can use the state as a mechanism to transfer wealth to themselves. In game theory, the prisoner’s dilemma and other non-cooperative games are essentially transaction cost problems. And other fields like industrial organization, international trade, development and labor, all contain ideas that hinge on the protection of property rights.

Given its long history and prevalence, it is ironic that the definition of transaction costs would be so difficult to agree on. This paper has argued that two definitions prevail in the literatures: one that defines transaction costs as only occurring when a market transaction takes place; the other defining
transaction costs as occurring whenever any property right is established or requires protection. I have called these the neoclassical and property rights definitions and have argued that which definition is useful depends on what question is being examined. Recognizing the distinction, though, is important for removing ambiguity and animosity.

Bibliography on Transaction Costs (0740)

Babic, Blagoje (1994), Prelaz u Tranziciji (Deception in Transition), Belgrade, Prometej.

Breenan, Michael J. (1975), ‘The Optimal Number of Securities in a Risky Asset Portfolio when there are Fixed Costs of Transaction: Theory and Some Empirical Results’, 10 Journal of Financial and Quantitative Analysis, 483-496.


Brennan, Michael J. (1975), ‘The Optimal Number of Securities in a Risky Asset Portfolio when there are Fixed Costs of Transaction: Theory and Some Empirical Results’, 10 Journal of Financial and Quantitative Analysis, 483-496.


Chianale, Angelo (1992), L’atto Pubblico: Contributo allo Studio dei Costi Transattivi (The Document under the Seal of a Public Officer: Contribution to the Study of Transaction Costs), Torino, Giappichelli.


West, E.G. (1990), Adam Smith and Modern Economics, Hants, Edward Elgar.