

the PAST and PRESENT of the THEORY of the FIRM[†]

Paul Walker[‡]

Abstract

In this survey we give a short overview of the way in which the theory of the firm has been formulated within the “mainstream” of economics, both past and present. As to a break point between the periods, 1970 is a convenient, if not entirely accurate, dividing line. The major difference between the theories of the past and the present, as they are conceived of here, is that the focus, in terms of the questions asked in the theory, of the post-1970 literature is markedly different from that of the earlier (neoclassical) mainstream theory. The questions the theory seeks to answer have changed from being about how the firm acts in the market, how it prices its outputs or how it combines its inputs, to questions about the firm’s existence, boundaries and internal organisation. That is, there has been a movement away from the theory of the firm being seen as developing a component of price theory, namely issues to do with firm behaviour, to the theory being concerned with the firm as a subject in its own right.

JEL Classification: B21, D23, L20.

Keywords: theory of the firm, neoclassical model, transaction costs, incomplete contracts.

28th November 2012

[†]Parts of this paper utilise material previously published in Walker (2010, forthcoming).

[‡]psw1937@gmail.com

Introduction

In this survey¹ we give a short overview of the way in which the theory of the firm² has been formulated within the “mainstream”³ of economics, both past and present.⁴ That there has been a close relationship between the general economic mainstream and the development of the theory of the firm has been noted by Foss and Klein (2006),

“[. . .] the evolution of the theory of the firm has never taken place far away from the economic mainstream. On the contrary, it has in fact been much driven by advances in the mainstream, and the relatively limited borrowing from other disciplines that has taken place has usually been strongly adapted to conform to central mainstream tenets. To be sure, the theory of the firm may have been revolutionary in the (some-what limited) sense of introducing new explananda to economics, but it is generally true to say that it has not been revolutionary in the sense of representing a radical break with any of the main tenets of mainstream economics”. (Foss and Klein 2006: 3-4).

Thus this survey’s concentration on the mainstream literature may do little damage to the story of the emergence of the theory of the firm but it does mean that little will be said of those non-mainstream or heterodox ideas, such as those from the overlap between economics and management or the Marxist approaches or the Austrian inspired theory of the firm or the relevant contributions from business history, that have developed outside of the orthodoxy.⁵

¹Parts of this essay utilise material previously published in Walker (2010, forthcoming).

²Spulber (2008: 5, footnote 8) gives the origin of the word ‘firm’ as “[t]he word ‘firm’ derives from the Latin word ‘firmare’ referring to a signature that confirmed an agreement by designating the name of the business”.

³Colander, Holt and Rosser (2004: 490) argue that the “[m]ainstream consists of the ideas that are held by those individuals who are dominant in the leading academic institutions, organizations, and journals at any given time, especially the leading graduate research institutions. Mainstream economics consists of the ideas that the elite in the profession finds acceptable, where by elite we mean the leading economists in the top graduate schools. It is not a term describing a historically determined school, but is instead a term describing the beliefs that are seen by the top schools and institutions in the profession as intellectually sound and worth working on”. Dequech (2007: 281) says “[. . .] that mainstream economics is that which is taught in the most prestigious universities and colleges, gets published in the most prestigious journals, receives funds from the most important research foundations, and wins the most prestigious awards”. In this survey we do not distinguish the “orthodoxy” from the “mainstream”. The terms are used interchangeably. See Colander, Holt and Rosser (2004, 2005) for a more sophisticated discussion of the concepts which draws a distinction between them.

⁴For much more complete surveys of the literature on the theory of the firm see, in chronological order, Boulding (1942), Papandeou (1952), Boulding (1960), Simon (1962), Cyert and March (1963: chapter 2), Machlup (1967), Cyert and Hedrick (1972), Williamson (1977), Milgrom and Roberts (1988), Tirole (1988: 15-61), Hart (1989), Holmstrom and Tirole (1989), Moore (1992), Borland and Garvey (1994), Hart (1995), Holmstrom and Roberts (1998), Foss (2000), Foss, Lando and Thomsen (2000), Khachatrian (2003), Garrouste (2004), Gibbons (2005), Mahoney (2005), Menard (2005), Garrouste and Saussier (2008), Müller (2009), Aghion and Holden (2011), Hart (2011b) and Zenger, Felin and Bigelow (2011). For surveys of the empirical literature see Joskow (1988), Shelanski and Klein (1995), Vannoni (2002), Klein (2005), Lafontaine and Slade (2007) and Hubbard (2008).

⁵Since the 1990s there has emerged a small Austrian literature on the firm, see for example Dulbecco and Garrouste (1999), Ioannides (1999), Witt (1999), Yu (1999), Lewin and Phelan (2000), Sautet (2000), Jankovic (2010) and Bylund (2011). For general discussions of this literature see Foss (1994, 1997), Foss and Klein (2009, 2010), Klein (2010) and Langlois (forthcoming). For discussions of the contributions from the resource-based theory of the firm see Wernerfelt (1984), Conner (1991), Lockett, O’Shea and Wright (2008) and Foss and Stieglitz (2010). On the knowledge-based view see Kogut and Zander (1992, 1996), Conner and Prahalad (1996) and Demsetz (1997). For critiques of knowledge-based theories see Foss (1996a, 1996b). Examples of the capabilities literature are Barney (1991) and Jacobides and Winter (2005). The most important work in the evolutionary economics approach to the firm is Nelson and Winter (1982). A discussion of the insightful but largely neglected paper, Malmgren (1961), is missing from this survey, but see Foss (1996c). For a discussion of some of the critics of the theory of the firm see Foss and Klein (2008). Sawyer (1979: chapter 9) considers ‘radical critique and radical alternatives’ to the theory of the firm. Sawyer briefly discusses Galbraith’s ‘theory of countervailing power’ (Galbraith 1963), Baran and Sweezy on *Monopoly Capital* (Baran and Sweezy 1966), Rothchild’s ‘Price Theory and Oligopoly’ (Rothchild 1947) and Galbraith’s *The New Industrial State* (Galbraith 1969). Hagendorf (2009) gives a Marxian critique of the theory of the competitive firm. The Marxian notion of the ‘conflict theory of the firm’ is examined in Baker and Weisbrot (1994). Another topic ignored here is the multinational firm, for an overview see Barba Navaretti et al (2004). On the relationship between the theory of the firm and entrepreneurship see Foss, Klein and Bylund (2011). For an overview of research into the growth of firms see Coad (2007, 2009). From business history comes Alfred D. Chandler’s classic works on the origins of the modern large-scale business enterprise, Chandler (1962, 1977, 1990). For a brief history of the development of the limited liability company see Hickson and

As to what constitutes the ‘past’ and the ‘present’, 1970 is a convenient, if not entirely accurate, dividing line since it was around this time that the present mainstream – largely Coaseian based – approaches to the firm started to develop with works such as Williamson (1971, 1973, 1975), Alchian and Demsetz (1972), Jensen and Meckling (1976) and Klein, Crawford and Alchian (1978). The major difference between the mainstream theories of the past and the mainstream theories of the present, at least as far as they are conceived of here, is that the focus – in terms of the questions the theory attempts to answer – of the post-1970 mainstream literature is markedly different from that of the earlier (neoclassical) mainstream theory. The theory of the firm for Ronald Coase, Oliver Williamson, Bengt Holmström or Oliver Hart is a very different thing from that of Arthur Pigou, Lionel Robbins, Jacob Viner, Joan Robinson or Edward Chamberlin. The questions the theory seeks to answer have changed from being about how the firm acts in the market, how it prices its outputs or how it combines its inputs, to questions about the firm’s existence, boundaries, including the dividing line between state and private enterprises, and internal organisation. That is, in the mainstream theory there has been a movement away from seeing the theory of the firm as simply developing one component (albeit an important component) of price theory, namely the element concerned with the factor and product market behaviour of producers, to the theory being concerned with the firm as an important economic institution in its own right.

The rest of this essay consists of four more sections. In Section 2, the past of the theory of the firm is considered. This section concentrates on a discussion of the classical theory of production and the neoclassical model of the ‘firm’. The third section of the paper consists of a short survey of the founding works – Knight (1921b) and Coase (1937) – on which the present versions of the mainstream theory of the firm are based while Section 4 deals with the current Coaseian/Knightian inspired theories themselves. Two general groups of theories are briefly discussed: principal-agent models and incomplete contract models. This section also contains a discussion of the theory of privatisation. Section 5 is a conclusion.

The ‘past’

Background⁶

While it can be argued that the theory of the firm has existed for only 80-90 years, in practice ‘firms’ have existed for several thousand years.⁷ Taking ancient India as an example, Khanna (2005) argues that the *sreni* - which was a complex organizational entity that shared similarities with corporations, guilds, and producers cooperatives - was being used as early as 800 B.C. and

Turner (2006). Walsh (2009) offers a Mengerian theory of the origins of the modern business firm. Another issue ignored here is corporate finance. On this see Zingales (2000) and Tirole (2006). The conditions under which different forms of firm ownership are optimal are discussed in Hansmann (1996).

⁶This section draws on material from Walker (2010).

⁷The first existence of a firm becomes especially problematic if we consider a farm to be a firm. Farming is an ancient human activity: “The first clear evidence for activities that can be recognized as farming is commonly identified by scholars as at about 12,000 years ago [. . .]”. (Barker 2006: 1). Tudge (1998: 3) writes “I want to argue that from at least 40,000 years ago – the late Palaeolithic – people were managing their environments to such an extent that they can properly be called ‘proto-farmers’”. At what historical point did the farm first become a firm? If we accept production for others as an important characteristic of the firm then farms can be seen (at least partially) as firms from a very early stage. Ofek (2001: chapter 13) argues that agriculture developed with a symbiotic relationship with exchange/trade. There is a conflict between the fact that we specialise in production but diversify in consumption. This conflict is reconciled by redistribution, i.e. via exchange/trade. Ridley (2010: 127-30) argues there would be no farming without trade, that trade was a precursor to farming: “One of the intriguing things about the first farming settlement is that they also seem to be trading towns. [. . .] it is a reasonable guess that one of the pressures to invent agriculture was to feed and profit from wealthy traders – to generate surplus that could be exchanged for obsidian, shells or other more perishable goods. Trade come first”. (Ridley 2010: 127). Spulber (2009: 103) takes a contrary position when he argues that the early farms were not firms. He writes that farms “from the earliest times to the eighteenth century are precursors to the contemporary firms. What distinguishes these economic actors from firms in that their enterprises tended to be integrated with the personal economic affairs of the entrepreneur. There was no separation between the owner’s commercial activities and their personal consumption activities”. For more on Spulber’s approach to the firm see pages 32–34 below. For a discussion of the economics of farms see Allen and Lueck (2002).

was in more or less continuous use from that time until 1000 A.D., at which time an Islamic invasion of India started.⁸ *Sreni* were utilised in occupations involving workers such as carpenters, ivory workers, bamboo workers, money-lenders, barbers, jewellers and weavers. (Khanna 2005: 10).⁹

In an analysis of the economy of the ancient Middle East Silver (1995: 50) notes that

“[p]rivate firms (*bītātu*) were prominent in late-third-millennium Akkad (the region south of Baghdad), in the Old Assyrian trade with Cappadocia [. . .] and, somewhat later, at Nippur. In the mid-second millennium the firm of *Tehip-tilla* played a major role in the real estate transactions and other business activities at Nuzi. A list of about the same time from Alalakh in northwest Syria refers to sixty-four firms participating in leatherworking, jewelry, and carpentry”.

Sobel (1999: 21) points out that during the Roman Republic contracting out of economic activities to private firms was the norm:

“[t]he republican Senate left virtually all economic activities to private individuals and companies, known collectively as the *publicani*. Tax collection, supplying the army, providing for religious sacrifices and ceremonies, building construction and repair, mining, and so on were all contracted out. There was even a contract for summoning the assembly in session and one for feeding the sacred geese”.

Micklethwait and Wooldridge (2003: 4) also note the private nature of tax collection in Rome, pointing out that companies were formed for this, and other purposes:¹⁰

⁸Importantly classical Islamic law does not grant standing to corporations, it recognises only national persons. For a discussion as to why Islamic law did not develop a concept akin to the corporation see Kuran (2005).

⁹*Sreni* were separate legal entities which could hold property separately from their owners, create their own regulations controlling the behaviour of their members, contract, sue and be sued in their own name. (Khanna 2005: 8-9). Khanna (2005: Table 1, p. 27; table footnotes removed) gives a summary of characteristics of the *sreni*:

Characteristics	Present in Ancient Indian <i>Sreni</i> ?
Separate Entity	Yes
Centralized Management	Yes
Transferability of Interest	Probably Yes
Limited Liability	Probably Not
Agent has power to bind entity?	Yes
Management elected?	Yes (though at times appears hereditary)
Can management be removed?	Yes
Duty of Loyalty Probably	Yes
Duty of Care	Yes
Liability insulation	Yes (though apparently not very detailed)
Screens on shareholder suits and internal enforcement activity	Yes (though apparently not very detailed)
Internal rules have binding effect	Yes
Some reimbursement for legal defense	Yes
Formation is easy	Yes
Register with state	Yes
State approval needed	Yes
Use of incentive payments	Yes (though apparently not very detailed)
Entry is easy	Some conditions, but no caste bars.
Sharing of assets and liabilities	Terms of agreement and additional rules
Exit is easy	Yes, but with obligations potentially
Board/Committee Independence	Probably Yes
Other board qualifications	Yes (though apparently not very detailed)
Voting Regulation	Yes (though apparently not very detailed)
Open debate in meetings & shareholder resolutions	Yes, with some limits (though apparently not very detailed)
Transparency is valuable and disclosure is encouraged	Probably Yes (though apparently not very detailed)

¹⁰For a brief discussion of the forms that firms could take in ancient Rome see Hansmann, Kraakman and Squire (2006: 1356-64).

“[t]he societates of Rome, particularly those organized by tax farming publicani, were slightly more ambitious affairs. To begin with, tax collecting was entrusted to individual Roman knights; but as the empire grew, the levies became more than any one noble could guarantee, and by the Second Punic War (218-202 b.c.), they began to form companies – *societates* – in which each partner had a share. These firms also found a role as the commercial arm of conquest, grinding out shields and swords for the legions. Lower down the social scale, craftsmen and merchants gathered together to form guilds (*collegia* or *corpora*) that elected their own managers and were supposed to be licensed”.

And some of these ancient firms were of reasonable size. Silver (1995: 66-7) notes,

“[w]e may note here that during the Ur III period a new mill at Girsu required the services of 679 women and 86 men (Maekawa 1980: 98)”

and

“[a] number of cities possessed large workshops employing hundreds of women in spinning and weaving. For example, a late-third-millennium text from Eshnunna lists 585 female and 105 male employees in a weaving house”. (Silver 1995: 143).

Ancient firms also diversified their activities.

“Large commercial houses flourished in Babylonia from the seventh to the fourth century. The House of Egibi, for example, bought and sold houses, fields, and slaves, took part in domestic and international trade, and participated in a wide variety of banking activities.

[:]

Earlier, in the late third-millennium Sumer, the rulers and governors controlled vertically integrated firms that used wool of the sheep they raised in their weaving workshops. At the same time, an Umma businessman (- bureaucrat?) named Ur-e-e busied himself with manifold operations, including raising livestock; transactions involving cheese, oil, leather, carcasses, wool; the weaving and finishing of cloth; shipments by boat of fish and grain; and even the construction of boats”. (Silver 1995: 67).

Thus the “firm” is an ancient and important empirical feature of the economic landscape but a feature which has been largely overlooked by economic theorists. The dichotomy between theory and practice could not be more stark. Even the present theories of the firm show that theorists have not long considered firms to be important economic entities. The current theories have been criticised as being rudimentary and bearing little relationship to the organisations we see in the world. As to the state of the modern theory of the firm Oliver Hart has written,¹¹

“[a]n outsider to the field of economics would probably take it for granted that economists have a highly developed theory of the firm. After all, firms are the engines of growth of modern capitalistic economies, and so economists must surely have fairly sophisticated views of how they behave. In fact, little could be further from the truth. Most formal models of the firm are extremely rudimentary, capable only of portraying hypothetical firms that bear little relation to the complex organizations we see in the world. Furthermore, theories that attempt to incorporate real world features of corporations, partnerships and the like often lack precision and rigor, and have therefore failed, by and large, to be accepted by the theoretical mainstream”. (Hart 1989: 1757).

¹¹In personal correspondence (November 2008 - used here with permission), Professor Hart said of the 1989 quote “[t]he language of 1989 is strong, and I’d probably tone it down a bit now. There’s been a lot of work in the last twenty years, and some progress. However, we are still not at the point where we have good models of the internal organization of large firms”.

This lack of an adequate theory of the firm, or an adequate theory of firm level production, is a problem of long standing in economics. More than 80 years before Hart, when surveying the history of the theory of production, Edwin Cannan argued that, “[b]efore the middle of the eighteenth century a theory of production can scarcely be said to have existed. Durable objects being looked upon as the sole or chief kind of wealth, the functions of industry and trade seemed to be the ‘circulation’ of wealth. When the physiocratic school turned the attention of economists to the consumable goods obtained by means of agriculture, the idea of circulation gave way to the idea of an annual reproduction, which gradually grew into the modern conception of production and consumption”. (Cannan 1903: 35-6). Cannan goes on to say that “‘Production’ and ‘Distribution’ do not seem, however, to have been used in England before 1821 as titles of divisions of political economy; and, before Adam Smith wrote, they were not in any sense technical economic terms”. (Cannan 1903: 32).

But the theories of production Cannan analysed were not theories of the firm, if we use the current mainstream approaches to modelling firms as the definition of such theories. It was noted by Cannan that “[o]ne of the most familiar and striking features of the theory of production, as taught in the text-books of the second half of the nineteenth century, is the practice of ascribing production to the co-operation or concurrence or joint use of three great agents, instruments, or requisites of production. Labour, Land, and Capital”. (Cannan 1903: 40). Such an approach has more in common with the later neoclassical production function approach to production¹² than the Coaseian inspired approaches utilised in the current mainstream theories of the firm.

The theories that Cannan was discussing were aimed at explaining the creation of the wealth of a nation¹³ rather than explaining the existence, boundaries and organisation of firms. Importantly it is not clear that the production which is being analysed in these theories is the production of wealth at the microeconomic level. Foss and Klein (2006: 7-8) note that classical economics was largely carried out at the aggregate level with microeconomic analysis acting as little more than a handmaiden to the macro-level investigation,

“[e]conomics began to a large extent in an aggregative mode, as witness, for example, the “Political Arithmetick” of Sir William Petty, and the dominant interest of most of the classical economists in distribution issues. Analysis of pricing, that is to say, analysis of a phenomenon on a lower level of analysis than distributional analysis, was to a large extent only a means to an end, namely to analyze the functional income distribution”.

The wealth considered in the production and distribution theories is either the total aggregate wealth of a nation or the average wealth of that nation and thus production is analysed at the macro level and not at the level of the firm. In his text “The Principles of Political Economy” Henry Sidgwick, for example, argued in terms of average wealth when he wrote that there were two fundamental questions that the theory of production attempts to answer: “(1) What are the causes that make the average annual produce per head of a given community at a given time greater than that of another whose primary wants are not materially different, or greater than its own produce at a previous stage of its history; and (2) What are the laws of their operation?” (Sidgwick 1901: 99). As late as 1935 Lionel Robbins could argue that the theory of production was about determining the total wealth, or total product, of the nation: “[t]he traditional approach to Economics, at any rate among English-speaking economists, has been by way of an enquiry into the causes determining the production and distribution of wealth. Economics has been divided into two main divisions, the theory of production and the theory of

¹²The first algebraic production function was most likely due to Johann Heinrich von Thünen in his book *The Isolated State*. (Humphrey 1997: 63-4). In a letter to Léon Walras dated January 6, 1877 Hermann Amstein derived the conditions of optimal factor hire from the competitive firm’s constrained cost function. He solved a cost-minimization problem in which the production function entered as a constraint. (Humphrey 1997: 66-8). Edgeworth (1889) uses a production function when giving the conditions for solving the profit maximisation problem. Production functions were common by the early 1890s, see for example, Berry (1891), Johnson (1891) and Wicksteed (1894).

¹³“‘Production’ and ‘distribution’ in political economy have always meant the production and distribution of wealth”. (Cannan 1903: 1). Chapter 1 of Cannan (1903) surveys the various meanings of “wealth” utilised in the economic writings of the 1776-1848 period.

distribution, and the task of these theories has been to explain the causes determining the size of the “total product” and the causes determining the proportions in which it is distributed between different factors of production and different persons”. (Robbins 1935: 64).

Blaug (1958: 226) went so far as to argue that the classical economists “[...] had no theory of the firm” and as will be argued below the neoclassical economists did little better in terms of a genuine theory of the firm.

This does raise the obvious question as to why economists ignored the firm, as an important economic institution in its own right, for so long.¹⁴ One reason for the neglect of the firm is simply that for a long time economists saw the internal workings of the firm to be outside the competence of economists. Arthur Pigou wrote:

“[...] it is not the business of economists to teach woollen manufacturers to make and sell wool, or brewers how to make and sell beer, or any other business men how to do their job. If that was what we were out for, we should, I imagine, immediately quit our desks and get somebody - doubtless at a heavy premium, for we should be thoroughly inefficient - to take us into his woollen mill or his brewery”. (Pigou 1922: 463-4).

Lord Robbins argued similarly, in that

“[t]he technical arts of production are simply to be grouped among the *given* factors influencing the relative scarcity of different economic goods. The technique of cotton manufacture [...] is no part of the subject-matter of Economics [...]”. (Robbins 1935: 33).

Foss and Klein (2006: 6-7) argue that there is the possibility of an empirical reason for the firm being overlooked; the relative unimportance of the firm. Until relatively recently firms were simply not a large part of the economy. But they also point out that such an explanation is not wholly convincing. Large firms¹⁵ have existed since at least the time of Adam Smith and the classical economists knew this. A more precise, and more defensible, version of the argument would be that the large, vertically integrated and diversified firm was not empirically important until

¹⁴As to why the firm was ignored in Austrian economics Witt (1999: 108) writes, “[t]he neglect of the firm as the organizational form of an entrepreneurial venture has a tradition in Austrian economics. It may be traced back to a characteristic of the scientific community in the German language countries. There, economic theory (Volkswirtschaftslehre) and business economics (Betriebswirtschaftslehre) were institutionally segregated as early as at the turn of the century to a degree still unknown today in the Anglo Saxon world. As Lachmann once conjectured, Austrian writers therefore considered the organizational form of entrepreneurial activities to be a topic best left to their business economics fellows”.

¹⁵Mokyr (2002: 122-3) summarises manufacturing in the U.K. before the Industrial Revolution by noting that “[...] large plants were not entirely unknown before the Industrial Revolution. For instance, Pollard (1968) in his classic work on the rise of the factory, mentions three large British plants, each employing more than 500 employees before 1750. Perhaps the most “modern” of all industries was silk throwing. The silk mills in Derby built by Thomas Lombe in 1718 employed 300 workers and were located in a five-story building. After Lombe’s patent expired, large mills patterned after his were built in other places as well. Equally famous was the Crowley ironworks, established in 1682 in Stourbridge in the Midlands (not far from Birmingham), which at its peak employed 800 employees. [...] In textiles, supervised workshops production could be found before 1770 in the Devon woollen industry and in calico printing (Chapman 1974)”. The development of factories and firms during the industrial revolution is discussed in Mokyr (2009: chapter 15). Also chartered companies were well known as witnessed by Adam Smith’s negative assessment of chartered companies in general and the East India Company in particular, contained in the *Wealth of Nations*. Jones and Ville (1996a: 898) note that “Adam Smith, no friend of chartered companies, argued that this separation of ownership from control contributed to gross administrative inefficiency, inattention to detail, and the pursuit of managerial goals, which raised prices to consumers and reduced returns to shareholders. He believed that only the extraction of monopoly rents ensured the success and continuance of such companies”. See Smith (1776: Book V, Chapter 1, Part e, pp. 731-58). Smith’s view of chartered companies is discussed in Kennedy (2010: 143-7). On the issue of whether the joint-stock chartered trading companies were an efficient institutional response to long-distance trade or were inefficient, rent-seeking monopolists see Carlos and Nicholas (1996) and Jones and Ville (1996a,b). A general history of the chartered companies is given in Cawston and Keane (1896), Griffiths (1974) and Ekelund and Tollison (1997: chapters 6 and 7). An important development for the modern large firm, following on from the chartered companies, was the introduction of limited liability. See Copp (2008) for a discussion of the reasons for the introduction of limited liability in the U.K. Limited liability protects investors from claims of the corporation, organisational law also does the converse. The assets of the corporation are protected from claims by investors. Hansmann, and Kraakman (2000a,b) and Hansmann, Kraakman and Squire (2005) emphasise the importance of this “asset separation” to the development of the firm. Hansmann, Kraakman and Squire (2006) traces the history of the emergence of entity shielding.

recently. Thus analysing anonymous “firms” may not have been a bad approximation to the empirical realities of the time.¹⁶ But the evidence presented above on the size and diversified nature of ancient firms as well as the size of some pre-industrial revolution firms (see footnote 15, p.6) should give us cause for reflection before accepting this conclusion without some reservations.

Neoclassical

For whatever reason it is only in more recent times that the firm has attracted attention as an important part of the economic system. As Foss, Lando and Thomsen (2000: 632) note:

“[i]t is only relatively recently, [. . .], that economists have felt the need for an economic theory addressing the reasons for the existence of the institution known as the (multi-person) business firm, its boundaries relative to the market, and its internal organization - to mention the issues that are generally seen as the main ones in the modern economics of organization [. . .]”.

Many would date the beginning of a genuine theory of the firm as recently as Knight (1921b) or Coase (1937), rather than to either the classical school or the neoclassical revolution.¹⁷ Before the contributions of Knight and Coase we had discussions of pin factories, but the discussion was about the importance of the division of labour rather than being ‘an enquiry into the nature and causes of the firm’.¹⁸ When discussing Adam Smith’s approach to the division of labour McNulty (1984: 237-8) comments,

“[h]aving conceptualized division of labor in terms of the organization of work within the enterprise, however, Smith subsequently failed to develop or even pursue systematically that line of analysis. His ideas on the division of labor could, for example, have

¹⁶As an approximation to “anonymous firm” production - that is, fully price-decentralised production - consider the case of rifle manufacture in Birmingham, England in the 1860s,

“[o]f the 5800 people engaged in this manufacture within the borough’s boundaries in 1861 the majority worked within a small district round St Mary’s Church. . . . The reason for the high degree of localization is not difficult to discover. The manufacture of guns, as of jewellery, was carried on by a large number of makers who specialized on particular processes, and this method of organization involved the frequent transport of parts from one workshop to another.

The master gun-maker-the entrepreneur-seldom possessed a factory or workshop. . . . Usually he owned merely a warehouse in the gun quarter, and his function was to acquire semi-finished parts and to give these out to specialized craftsmen, who undertook the assembly and finishing of the gun. He purchased materials from the barrel-makers, lock-makers, sight-stampers, trigger-makers, ramrod-forgers, gun-furniture makers, and, if he were engaged in the military branch, from bayonet-forgers. All of these were independent manufacturers executing the orders of several master gun-makers. . . . Once the parts had been purchased from the “material-makers,” as they were called, the next task was to hand them out to a long succession of “setters-up,” each of whom performed a specific operation in connection with the assembly and finishing of the gun. To name only a few, there were those who pre-pared the front sight and lump end of the barrels; the jiggers, who attended to the breech end; the stockers, who let in the barrel and lock and shaped the stock; the barrel-strippers, who prepared the gun for rifling and proof; the hardeners, polishers, borers and riflers, engravers, browners, and finally the lock-freers, who adjusted the working parts”. (Allen (1929: 56-7 and 116-7), quoted in Stigler (1951: 192-3).)

Such a method of production would be a guide to the way production would take place under a functioning version the neoclassical model of the “firm”. It could be argued that this form of production isn’t neoclassical since it is not clear that the neoclassical separation theorem is satisfied. See Spulber (2009) for a discussion of the separation theorem.

¹⁷O’Brien (1984: 25) takes a contrary position: “[s]erious discussion of the history of the theory of the firm has to start with Alfred Marshall”. O’Brien’s argument is based, in the main, on Marshall (1920). O’Brien also argues that developments subsequent to Marshall have resulted in many of Marshall’s insights being lost to succeeding generations of economists. We would therefore argue that Marshall has left little in the way of a legacy in terms of the mainstream theory of the firm. In addition to his views on Marshall’s work and later developments O’Brien also argues that any “attempt to construct a pre-Marshallian theory from the materials available is likely to be unsuccessful”. See, however, Williams (1978) for such an attempt. On the neglect of Marshall’s ‘Industry and Trade’ (Marshall 1920) see also Liebhafsky (1955). The development of the “theory of the firm” from Marshall to Robinson and Chamberlin is also dealt with in Moss (1984).

¹⁸When writing about Adam Smith’s approach to the firm Williams (1978: 11) says, “[t]he firm was disembodied and became a unit in which resources congeal in the productive process. When we come to examine the equilibrium/value theory of *The Wealth of Nations* it will be shown that, in that context, the firm is little more than a passive conduit which assists in the movement of resources between alternative activities.”

led him towards an analysis of task assignment, management, or organization". Such an approach would have foreshadowed the much later—indeed, quite recent—effects in this direction by Herbert Simon, Oliver Williamson, Harvey Leibenstein, and others, a body of work which Leibenstein calls “micro-micro economics”. [...] But, instead, Smith quickly turned his attention away from the internal organization of the enterprise, and outward toward the market and the realm of exchange, perhaps because he found therein both the source of division of labour, in the “propensity in human nature [...] to truck, barter and exchange” and its effective limits”.

As has been pointed out by Demsetz (1982, 1988a, 1995) before Knight and Coase – and it could be added for much of the period after them – the fundamental preoccupation of economists was with the market and the price system and hence little, or no, attention was paid to either the firm or the consumer as separate, significant, economic entities. Firms (and consumers) existed as handmaidens to the price system.

The interest in the price system, culminating in the “perfect competition” model, has its intellectual origins in the eighteenth-century debate between free traders and mercantilists. Butler (2007: 25-6) briefly sums up mercantilism in the following way:¹⁹

“[...] it measured national wealth in terms of a country’s stock of gold and silver. Importing goods from abroad was seen as damaging because it meant that this supposed wealth must be given up to pay for them; exporting goods was seen as good because these precious metals came back. Trade benefited only the seller, not the buyer; and one nation could get richer only if others got poorer. On the basis of this view, a vast edifice of controls was erected in order to prevent the nation’s wealth draining away - taxes on imports, subsidies to exporters and protection for domestic industries. [...] Indeed, all commerce was looked upon with suspicion and the culture of protectionism pervaded the domestic economy too. Cities prevented artisans from other towns moving in to ply their trade; manufacturers and merchants petitioned the king for protective monopolies; labour saving devices such as the new stocking-frame were banned as a threat to existing producers”.

The free trade versus mercantilism debate was, to a large degree, about the proper scope of government in the economy²⁰ and the model it (eventually) gave rise to reflects this. The question implicitly at the centre of the debate was, Is central planning necessary to avoid the problems of a chaotic economic system? Adam Smith famously answered “no”.²¹ Smith

¹⁹For a detailed discussion of mercantilism see Heckscher (1934), Viner (1937), Magnusson (1994) and Ekelund and Tollison (1997).

²⁰Mercantilism requires a dominate state, to provide and enforce monopolies and to regulate and control domestic and international trade and direct the economy in general. “Mercantilist policies include the use of state power to build up industry, to obtain and increase the surplus of exports over imports, and to accumulate stocks of precious metals”. (Blackhouse 2002: 58). “In France during this period [mid-1700s] the concept [mercantilism] was utilized in order to describe an economic policy regime characterized by direct state intervention, intended to protect domestic merchants and manufacturers”. (Magnusson 2003: 46). When discussing the general economic background to the development of the mercantile chartered companies in England, Griffiths (1974) explains that “[t]he right—and the duty—of the Crown to control the economy was taken for granted and according to Coke ‘the royal prerogative had an ancient and special force in the government of trade’” (p. ix) and “[t]he underlying concepts were those of monopolies, collective trading or regulation of trade and the right of the Crown to control the economy”. (p. 3). In a comment on Eli Heckscher’s view of mercantilism Deepak Lal writes that “Heckscher had argued that the mercantilist system arose as the Renaissance princes sought to consolidate the weak states they had inherited or acquired from the ruins of the Roman Empire. These were states encompassing numerous feuding and disorderly groups which the new Renaissance princes sought to curb to create a nation. The purpose was to achieve “unification and power,” making the “State’s purposes decisive in a uniform economic sphere and to make all economic activity subservient to considerations corresponding to the requirements of the State.” The mercantilist policies—with their industrial regulations, state-created monopolies, import and export restrictions, price controls—were partly motivated by the objective of granting royal favors in exchange for revenue to meet the chronic fiscal crisis of the state [...]. Another objective was to extend the span of government control over the economy to facilitate its integration”. (Lal 2006: 307).

²¹According to Smith the government has three duties: “[t]he first duty of the sovereign, that of protecting the society from the violence and invasion of other independent societies [...]”. (Smith 1776: Book V, Chapter 1, Part First, p. 689). “The second duty of the sovereign, that of protecting, as far as possible, every member of the society from injustice or oppression of every other member of it, or the duty of establishing an exact administration of justice, [...]”. (Smith

“[. . .] realised that social harmony would emerge naturally as human beings struggled to find ways to live and work with each other. Freedom and self-interest need not lead to chaos, but – as if guided by an ‘invisible hand’ – would produce order and concord. They would also bring about the most efficient possible use of resources. As free people struck bargains with others – solely in order to better their own condition – the nation’s land, capital, skills, knowledge, time, enterprise and inventiveness would be drawn automatically and inevitably to the ends and purposes that people valued most highly. Thus the maintenance of a prospering social order did not require the continued supervision of kings and ministers. It would grow organically as a product of human nature”. (Butler 2007: 27-8).

For Smith competitive markets were the most prominent mechanism for coordinating and motivating people to maximise the gains that result from increased specialisation and an expanded division of labour. Well functioning market institutions leave individuals free to pursue self-interested behaviour, but guide their choices by the prices they pay and receive. For economists, the 200 years following Smith involved a search for conditions under which the price system would function well, conditions under which it would not descend into chaos.

The formal (neoclassical) model that arose from this search is one which abstracts completely from any form of centralised control in the economy.²² It is a model delineated by “perfect decentralisation”.²³ Authority, be it in the form of a government or a firm or a household, plays no role in coordinating resources.²⁴ The only parameters guiding decision making are those given within the model – tastes and technologies – and those determined impersonally on markets –

1776: Book V, Chapter 1, Part II, p. 709). “The third and last duty of the sovereign or commonwealth is that of erecting and maintaining those publick institutions and those publick works, which, though they may be in the highest degree advantageous to a great society, are, however, of such a nature that the profit could never repay the expense to any individual or small number of individuals, and which it therefore cannot be expected that any individual or small number of individuals should erect or maintain”. (Smith 1776: Book V, Chapter 1, Part III, p. 723). For book length discussions of Smith’s thought see, for example, Evensky (2005), Kennedy (2005, 2010) and Otteson (2002, 2011).

²²For Adam Smith this would be an abstraction too far. Smith knew of the importance of institutions to the proper functioning of the market economy. Mark Blaug points out that “[. . .] Smith’s faith in the benefits of ‘the invisible hand’ has absolutely nothing whatever to do with allocative efficiency in circumstances where competition is perfect à la Walras and Pareto; the effort in modern textbooks to enlist Adam Smith in support of what is now known as the ‘fundamental theorems of welfare economics’ is a historical travesty of major proportions. For one thing, Smith’s conception of competition was, as we have seen, a process conception, not an end-state conception. For another society, a decentralised competitive price system was held to be desirable because of its dynamic effects in widening the scope of the market and extending the advantages of the division of labour - in short, because it was a powerful engine for promoting the accumulation of capital and the growth of income”. (Blaug 1996: 60-1).

²³The neoclassical model is often described as one of “perfect competition” and one reason that the emphasis on the firm diminished as the model developed was that the neoclassical placed a growing emphasis on the concept of market competition and thus less emphasis was given to the firm. As McNulty (1984: 240) explains “[t]he ‘perfection’ of the concept of competition, beginning with the work of A. A. Cournot and ending with that of Frank Knight, which was at the heart of the development of economics as a science during the nineteenth and early twentieth centuries, led on the one hand to an increasingly rigorous analytical treatment of market processes and on the other hand to an increasingly passive role for the firm”. For Knight “[p]erfect competition is conditioned by the existence of a set of assumptions, the most important of which are the following: (1) “a perfect market for productive services [. . .], that is, uniform prices over the whole field” (1921[a], 316); (2) complete rationality and perfect knowledge by free and independent individuals; (3) “perfect mobility in all economic adjustments, no cost involved in movements or changes” (1921[b], 77); (4) “virtually instantaneous and costless” exchange of commodities (1921[b],78); (5) “perfect, continuous, costless intercommunication between all individual members of the society” (1921[b], 78); (6) perfect divisibility of commodities; and (7) “an indefinitely large number of competing organizations, each of the most efficient size” (1921[a], 316)”. (Marchionatti 2003: 58).

²⁴The household in the neoclassical model is as lacking in substance as the firm. Kenneth Boulding made the point that “[t]his type of analysis [the theory of the firm] is exactly analogous to the analysis of the reactions of a consumer by means of indifference curves. Indeed, a consumer is merely a “firm” whose product is “utility.” The indifference curves are analogous to the isoquants, or product contours, the only difference being that they cannot be assigned definite quantities of utility. The utility surface, whose contours form the system of indifference curves, is a “mountain” whose shape we theoretically know, but whose height at any point probably cannot be known; by contrast, we can assume that both shape and height of the production surface are known. The “substitution effect” and the “scale effect” are likewise known in consumption theory, where the scale effect is usually called the “income effect.” Thus, a rise in the price of a single object of consumption will have a substitution effect tending to reduce the consumption of that object as cheaper alternatives are substituted for it. There will also be an “income effect” tending to reduce all consumption, as the higher price makes the consumer poorer. The effect of a given rise in price, therefore-i.e., the elasticity of demand-depends first on the substitutability of the commodity concerned, and, secondly, on its importance in the total expenditure. This is true either

prices. All parameters are outside the control of any of the economic agents and this effectively deprives all forms of authority a role in allocation. This includes, of course, the firm. It doesn't matter whether it is the general equilibrium version of the neoclassical model, characterised by Walras's *tâtonnement* process, or the partial equilibrium version, characterised by Pigou's equilibrium firm, there is no serious consideration given to the firm as a problem solving institution.²⁵

Like so much of neoclassical economics it was Alfred Marshall who gave us much of the foundations of the neoclassical theory of the firm. It was Marshall's notion of the "representative firm" that evolved into the now common textbook theory of the firm. For Marshall firms were dynamic, they progressed through a life cycle in much the same way as people. "They began young and vigorous, but after a period of maturity they became old and were displaced by newer more efficient firms". (Backhouse 2002: 179). Marshall gave us the famous metaphor of an industry being like a forest—while it might appear unchanged if considered as a whole, the individual trees that make it up are constantly changing. To reconcile his dynamic view of individual firms with the static view of industries Marshall introduced his idea of a "representative firm". The representative firm was "composed of the salient characteristics of all firms in the industry". (Moss 1984: 308). For Marshall his analysis of the firm sort to rationalise his studies of real world firms while the idea of the industry was an abstract concept under the umbrella of which the various producers of goods and services could be grouped to facilitate the analysis of the problem at hand. The role of the representative firm was to link the dynamic view of the firm with the abstract view of the industry.²⁶

Moss (1984a,b) argues that there were three crucial steps in the movement from the Marshallian to the now 'textbook' view of the firm. The first step began with the publication of "Wealth

of a consumption good or of a factor of production". (Boulding 1942: 799). Fritz Machlup argues that the household is not the subject of study in the theory of the consumer: "[t]he "household" in price theory is not an object of study; it serves only as a theoretical link between changes in prices and changes in labor services supplied and in consumer goods demanded. The hypothetical reactions of an imaginary decision-maker on the basis of assumed, internally consistent preference functions serve as the simplest and heuristically satisfactory explanation of empirical relationships between changes in prices and changes in quantities. In other words, the household in price theory is not an object of study". (Machlup 1967, footnote 4, p. 9).

²⁵About the partial equilibrium approach to the firm Klein (1996: 5) writes,

"[i]n neoclassical economic theory, the firm as such does not exist at all. The "firm" is a production function or production possibilities set, a means of transforming inputs into outputs. Given the available technology, a vector of input prices, and a demand schedule, the firm maximizes money profits subject to the constraint that its production plans must be technologically feasible. That is all there is to it. The firm is modeled as a single actor, facing a series of relatively uncomplicated decisions: what level of output to produce, how much of each factor to hire, and so on. These "decisions," of course, are not really decisions at all; they are trivial mathematical calculations, implicit in the underlying data. In the long run, the firm may also choose an optimal size and output mix, but even these are determined by the characteristics of the production function (economies of scale, scope, and sequence). In short: the firm is a set of cost curves, and the "theory of the firm" is a calculus problem".

The high water mark for neoclassical general equilibrium approach is arguably Debreu (1959). For Debreu there are no firms, in the normal sense of the word, there are just "producers",

"[...] when one abstracts from legal forms of organization (corporations, sole proprietorships, partnerships, ...) and types of activity (Agriculture, Mining, Construction, Manufacturing, Transportation, Services, ...) one obtains the concept of a producer, i.e., an economic agent whose role is to choose (and carry out) a production plan". (Debreu 1959: 37).

It is also clear from the context that the agent referred to is a person. The only role for the agent is to pick the profit maximising production plan from the set of available plans. Langlois (1981: 5) explains that "[...] the interesting feature of the general-equilibrium formulation is not so much that it takes as given the mix of market and internal transactions; rather, it is that the assumptions of general-equilibrium theory themselves actually suggest that there need be no internal activity whatsoever. If all commodities are predetermined for all time and the techniques for producing them are given and fully known in all details, then one could easily conceive of a situation where every separate part of the production process would be in the nature of a market transaction".

²⁶Hart (2003: 1140) writes,

"[i]t [the representative firm] was an avenue through which Marshall conjectured a notion of equilibrium at a point in time for the industry as a whole, while at the same time individual firms were in disequilibrium, being subject to an "organic" process of change. The Representative Firm therefore meets at the junction of Marshall's biological and mechanical notions of opposed forces described in the introductory comments in book 4 of Principles".

and Welfare” by A. C. Pigou. Pigou utilised a formalisation of Marshall’s industrial taxonomy, that is the distinction between constant, increasing and decreasing returns to scale industries, to study the effect of these industries on the national dividend. Pigou applied this taxonomy in “Wealth and Welfare” and the first edition of “The Economics of Welfare” in a wholly abstract manner and this abstract analysis was the target of Clapham’s (1922) attack on Marshall and Pigou. Clapham argued that it was not in general possible to assign actual industries to any one of the three categories. In addition Sraffa (1926) raised two objections to Marshall’s categorisation. First Sraffa argued that increasing returns to scale were incompatible with perfect competition. Assuming external increasing returns to scale meant reliance on a class of returns that were “seldom to be met with” (Sraffa 1926: 540). That it is impossible to reconcile decreasing returns to scale with perfect competition is Sraffa’s second objection. As many firms incurred some form of fixed costs, it follows that they would be producing subject to diminishing costs (or increasing returns). The problem for the static theory of the industry is that a firm that faces a given price and produces under increasing returns to scale will increase its output without limit. Thus increasing returns were incompatible with perfect competition. As part of a response to Clapham’s claim of empirical irrelevance and Sraffa’s claim of logical incoherence Pigou developed further Marshall’s notion of increasing returns to scale which are internal to the industry but external to the firm.

This development is the second of Moss’s three steps and, importantly, involved Pigou introducing the “equilibrium firm”.²⁷ In contrast to the representative firm, which can be in disequilibrium when the industry is in equilibrium, an equilibrium firm would be in equilibrium whenever the industry is in equilibrium. In his 1928 paper ‘An Analysis of Supply’ Pigou outlined the conditions for equilibrium which, importantly, involves all the internal economies of scale being exhausted, that is, the output level of the equilibrium firm, for a many-firm industry, would occur where the marginal cost curve cuts the average cost curve. (Pigou 1928: 254).

The last of the three steps was to assume that industries are comprised entirely of equilibrium firms with identical cost curves, and to assume that firms, as production functions²⁸, faced house-

²⁷In Pigou (1928: 239-40) he describes the equilibrium firm, at some length, as

“[m]ost industries are made up of a number of firms, of which at any moment some are expanding, while others are declining. Marshall, it will be remembered, likens them to trees in a forest. Thus, even when the conditions of demand are constant and the output of an industry as a whole is correspondingly constant, the output of many individual firms will not be constant. The industry as a whole will be in a state of equilibrium; the tendencies to expand and contract on the part of the individual firms will cancel out; but it is certain that many individual firms will not themselves be in equilibrium and possible that none will be. When conditions of demand have changed and the necessary adjustments have been made, the industry as a whole will, we may suppose, once more be in equilibrium, with a different output and, perhaps, a different normal supply price; but, again, many, perhaps all, the firms contained in it, though their tendencies to expand and contract must cancel one another, will, as individuals, be out of equilibrium. This is evidently a state of things the direct study of which would be highly complicated. Fortunately, however, there is a way round. Since, when the output of the industry as a whole is adjusted to any given state of demand, the tendencies to expansion and contraction on the part of individual firms cancel out, they may properly be regarded as irrelevant so far as the supply schedule of the industry as a whole is concerned. When the conditions of demand change, the output and the supply price of the industry as a whole must change in exactly the same way as they would do if, both in the original and in the new state of demand, all the firms contained in it were individually in equilibrium. This fact gives warrant for the conception of what I shall call the *equilibrium firm*. It implies that there *can* exist some one firm, which, whenever the industry as a whole is in equilibrium, in the sense that it is producing a regular output y in response to a normal supply price p , will itself also individually be in equilibrium with a regular output x_r . The conditions of the industry are compatible with the existence of such a firm; and the implications about these conditions, which, whether it in fact exists or not, would hold good if it did exist, must be valid. For the purpose of studying these conditions, therefore, it is legitimate to speak of it as actually existing. For any given output, then, of the industry as a whole, the supply price of the industry as a whole must be equal to the price, which, with the then output of the industry as a whole, leaves the equilibrium firm in equilibrium. The industry, therefore, conforms to the law of increasing, constant or decreasing supply prices according as the price which leaves the equilibrium firm in equilibrium increases, remains constant, or decreases with increases in the output of the industry as a whole”.

²⁸Moss (1984a: 313) notes that Pigou’s analysis of the equilibrium firms gave us the firm as a production function. “Whatever its relationship to the representative firm, Pigou’s introduction of the equilibrium firm gave us the firm as production function. In “An analysis of Supply”, Pigou demonstrated diagrammatically the various possible relationships between marginal and average curves on the assumption, made clear in his algebraic analysis, that factor prices were

hold preference functions. This task was carried out by Robinson (1933) and Chamberlin (1933) in their development of imperfect competition and monopolistic competition, respectively.²⁹ But as Moss (1984a: 314) points out,

“[b]y assuming that every firm in the industry has an identical cost curve, Robinson and Chamberlin stood Pigou’s construction of the equilibrium firm on its head. Where Pigou argued that an equilibrium firm could be derived from the laws of returns obeyed by any particular industry, Robinson and Chamberlin defined the industry on the basis of a population of equilibrium firms”.

Thus, by the 1930s the neoclassical approach³⁰ to the firm had developed. But many economists would argue that the neoclassical model isn’t a “theory of the firm” in any meaningful sense.³¹ The output side of the standard neoclassical model is a theory of supply or production rather than a true theory of the firm. In neoclassical theory, the firm is a ‘black box’ there to explain how changes in inputs lead to changes in outputs.³² The firm is a conceptualisation that represents, formally, the actions of the owners of inputs who place their inputs in the highest value uses, and makes sure that production is separated from consumption. The firm produces only for outsiders, there is no on-the-job or internal consumption, no self-sufficiency. In fact there are no managers or employees to indulge in on the job consumption and as production is separated from consumption, no self-sufficiency. Production for outsiders is, according to Demsetz (1995), the definition of a firm in the neoclassical model:

“[w]hat is needed is a concept of the firm in which production is exclusively for sale to those formally outside the firm. This requirement defines the firm (for neoclassical theory), but it has little to do with the management of some by others. The firm in neoclassical theory is no more or less than a specialized unit of production, but it can be a one-person unit”. (Demsetz 1995: 9).

As inputs are combined in the optimal fashion by the actions of independent input owners motivated solely by market prices, there is no need for ‘management of some by others’, there is

either unchanged or compensated. All changes in average and marginal costs were due to technological factors alone, and since the equilibrium firm was characterized by given average and marginal cost curves which did not shift as a result of any activity of the firm, those technological factors were considered to be entirely exogenous to the firm. Pigou’s technique here was analytically equivalent to the derivation of a cost curve from the expansion path of a production function”.

²⁹Backhouse (2003: 315) writes that Robinson’s *Economics of Imperfect Competition* “[...] virtually created the modern geometry of the theory of the firm, analyzing perfect and imperfect competition, monopoly, monopsony, and even the kinked demand curve (conventionally attributed to Sweezy, 1939)”. Shackle (1967: 61-2) writes, “[t]he two books [Robinson’s and Chamberlin’s] are very different in scope. Mrs Robinson’s central concern is with the effect of supposing the demand for firm’s output to be less than perfectly elastic, so that each firm, though only one among a multitude of firms producing substitutes of varying closeness for each other’s products, can exploit the essential position, powers and policies of a monopolist. She eschews discussion of those markets where each firm reckons on other firms’ active retaliation to its moves, and of expenditure on selling effort. [...] Professor Chamberlin includes selling expenditure in his analysis with a most ingenious formal precision. He also duplicates many arguments about price behaviour in order to point out that the entrepreneur should consider the profit possibilities of all products and choose in the end that output of that product which, with the optimal selling expenditure, yields the biggest total profit. With these ostensibly large extensions of the field, compared with Mrs Robinson’s; with different emphases and a chief reliance on different diagrammatic tools; and especially with a personal interpretation of such words as ‘supply’ and with impalpable distinctions between his own and Mrs Robinson’s use of the expressions ‘monopoly’, ‘imperfect competition’ and others, Professor Chamberlin is at great pains to insist that the two approaches are essentially different. Almost all other students of the matter have agreed with each other that in describing the structure and mechanism of equilibrium in firms and groups of firms when oligopoly and selling expenditure are absent, the two books present identical theories”. Chamberlin (1937) discusses the differences between monopolistic and imperfect competition.

³⁰Or the marginalist theory of the firm as it was often referred to at this time. Mongin (1997: 558) notes that marginalist was the commonly used term in the 1940s and 50s, as “the term neo-classical was not yet popular”. The term marginalist was still being used in the 1960s as witnessed by the title of Fritz Machlup’s 1967 American Economic Association Presidential Address, ‘Theories of the Firm: Marginalist, Behavioral, Managerial’. (Machlup 1967).

³¹For an example of an approximation to what production could look like in a neoclassical world see footnote 16, p.7.

³²It is a black box in the sense that inputs go in and outputs come out, without any explanation of how one gets turned into the other. The firm is taken as given; no attention is paid to how it came into existence, the nature of its internal organisation, where the boundary between one firm and another is or between a firm and the market; or whether anything would change if two firms merged and called themselves a single firm.

no role for managers or employees. Also note that as competition assures the absence of profits and losses in equilibrium, there is no need to have a residual claimant. This means that, in one sense at least, there are no owners of the firm.³³ As there are no physical assets controlled by the firm, there are no (residual) control rights over these assets to allocate. This implies there are no owners of the firm in the Grossman-Hart-Moore sense.

The neoclassical production function is a way of representing the (efficient) black box conversion of inputs into outputs but tells us little about the inner workings of the black box. The production function is independent of the institutional framework of output creation. It can be given, at least, two interpretations: it can represent the production method of a single firm, of which all known firms are just divisions or, equally, it could represent the outcome of a series of purely market based transactions which give rise to the observed outputs.

Thus it represents the 'firm' without explaining the 'firm'. The boundaries of the firm is an issue described by Williamson (1993: 4) as one of

“[. . .] make-or-buy. What is it that determines which transactions are executed how? That posed a deep puzzle for which the firm-as-production function approach had little to contribute”.

Hart (1995: 17) criticises the neoclassical model based on three characteristics of the theory. First, he notes that the theory completely ignores incentive problems within the firm. The firm is a perfectly efficient 'black box'. Second, the theory has nothing to say about the internal organisation of the firm. Nothing is said about the hierarchical structure, how decisions are made, who has authority within a firm. Third, the theory tells us nothing about how to pin down the boundaries of the firm. The theory is as much a theory of plant or division size as firm size. As Hart points out,

“[t]o put it in stark terms [. . .] neoclassical theory is consistent with there being one huge firm in the world, with every existing firm [. . .] being a division of this firm. It is also consistent with every plant and division of an existing firm becoming a separate and independent firm”. (Hart 1995: 17).

While the neoclassical model is consistent with both there being only one huge firm in the world and with there being a very large number of very small firms, as Foss (2000) notes the model is also consistent with there being no firms, since consumers can do it all!

“With perfect and costless contracting, it is hard to see room for anything resembling firms (even one-person firms), since consumers could contract directly with owners of factor services and wouldn't need the services of the intermediaries known as firms”. (Foss 2000: xxiv)

Cyert and Hedrick (1972) addressed similar points. They argue that in the neoclassical system the firm doesn't exist, that no real world problems of firms are considered, that there are no organisational problems or any internal decision-making process at all.

“In one sense the controversy over the theory of the firm has arisen over a non-existent entity. The crux of microeconomics is the competitive system. Within the competitive model there is a hypothetical construct called the firm. This construct consists of a single decision criterion and an ability to get information from an external world, called the “market” [8, Cyert and March, 1963, pp. 4-16]. The information received from the market enables the firm to apply its decision criterion, and the competitive system then proceeds to allocate resources and produce output. The market information determines the behavior of the so called firm. None of the problems of real firms

³³Hansmann (1996), for example, states “[a] firm's “owners,” as the term is conventionally used and as it will be used here, are those persons who share two formal rights: the right to control the firm and the right to appropriate the firm's profits, or residual earnings (that is, the net earnings that remain with the firm after it has made all payments to which it is contractually committed, such as wages, interest payments, and prices for supplies)”. (p. 11) He later adds “[n]ot all firms have owners. In nonprofit firms, in particular, the persons who have control are barred from receiving residual earnings”. (p. 12).

can find a home within this special construct. There are no organizational problems nor is there any room for analysis of the internal decision-making process". (Cyert and Hedrick 1972: 398).

Thus within the neoclassical model of the price system, the firm's only role is to allow input owners to convert inputs into outputs in response to market prices. Firms have no internal organisation since they have no need of one, they have no owners since there is nothing to own. Questions about the existence, definition, internal structure and boundaries of the firm are to a large degree meaningless within this framework since firms, by any meaningful definition of that term, do not exist. As Foss, Lando and Thomsen (2000: 632) summarise it:

"[t]he pure analysis of the market institution leaves almost no room for the firm (Debreu 1959). Under the assumption of a perfect set of contingent markets, as well as certain other restrictive assumptions, the model describes how markets may produce efficient outcomes. The question how organizations should be structured does not arise, because market-contracting perfectly solves all incentive and coordination issues. By assumption, firm behaviour (profit maximization) is invariant to institutional form (e.g. ownership structure). The whole economy can operate efficiently as one great system of markets, in which autonomous agents enter into very elaborate contracts with each other. However, by treating the firm itself as a black box, where internal structure, contracts, etc. disappear from the picture, there are many other issues that the theory cannot address. For example, the theory does not tell us why firms exist".

Despite the fact that by the 1930s the neoclassical approach was the dominate theory of the firm, the exact role of the theory of the firm in price theory was still the subject of much confusion and debate. The most famous of these debates were the "full cost controversy" and the related "marginalist controversy" (Mongin 1992, 1998). The full cost controversy was started by the publication in 1939 of a paper by R. L. Hall and C. J. Hitch which looked at pricing policies of firms (Hall and Hitch 1939). On the basis of questionnaire data Hall and Hitch argued that firms set prices in a "full-cost" way by estimating an average-cost amount at a reference level of output and adding to it a fixed percentage. Full-cost pricing came to be seen as a challenge to the usual marginalist (neoclassical) profit-maximising view of the firm. Long-run profit maximisation would only be achieved if the mark-up bore the correct relationship to the firm's perceived elasticities of demand. The most famous defense of the marginalist theory came from Machlup (1946). Machlup's response, however, wasn't solely directed towards the full-cost arguments, he also attacked a paper by labour economist R. A. Lester which argued that the theoretical predicts regarding the relationship between wages and employment could not be found in the data (Lester 1946). Lester argued that "[...] his empirical research raised "grave doubts as to the validity of conventional marginal theory and the assumptions on which it rests" in the following ways: (1) market demand was more important in determining a firm's volume of employment than wage rates; (2) the firm's cost structure was not that suggested by "conventional marginalism" and its capital-labor ratio was not tied to its wage rate structure; and (3) "the practical problems involved in applying marginal analysis to the multi-process operations of a modern plant seem insuperable, and business executives rightly consider marginalism impractical as an operating principle in such manufacturing establishments" [Lester 1946, pp. 81-82]" (Lee 1984: 1114). Lester's conclusion was that businessman did not adjust their employment levels in relationship to changes in wages and productivity in a manner consistent with the marginal theory.

At this point it is difficult to separate out the full-cost controversy from the Lester initiated marginalist controversy. In reply to both sets of arguments Machlup "[...] managed to dispute the quality and relevance of the evidence, and at the same time, to claim that data on price-setting were compatible with several of the available models of imperfect competition; he also sketched a general decision-theoretic argument to the effect that "rules of thumb" (the expression in Hall and Hitch) often reflect an underlying optimizing process. Most of the later neoclassical arguments are already in Machlup's proteistic plea. His general conclusion was that the current

theory of the firm hardly needed revising even if the allegedly damaging findings were taken at face value” (Mongin 1992: 314-5).

Effectively these controversies ended when Richard B. Heflebower presented a paper at the Conference on Business Concentration and Price Policy in June 1952 (Heflebower 1955). Heflebower showed that full-cost pricing could be viewed in marginalist terms. He argued that profit maximisation should be understood in a long-run sense and that oligopoly should become the main theoretical focus for economists. He added that the full-cost doctrine did not constitute a well developed body of price theory and that the empirical work on which it was based was “spotty in quality”.

Importantly little changed because of these controversies. As Mongin (1998: 280) notes, for the majority of economists “[...] drastic adjustments in the theory of the firm were not needed to resolve the marginalist controversy”. Overall, “[a]lthough no contribution to the AER controversy [the marginalist controversy] can be said to be decisive, it can be conjectured that it influenced American economists into thinking that Robinsons and Chamberlins initial models had to be refined, but that the profit-maximizing framework was flexible enough to accommodate the available evidence” (Mongin 1998: 279) and “[i]t is clear from Heflebower’s masterly survey that many of the arguments used by supporters of the fullcost principle are in no way inconsistent with orthodox economic theory”. (Coase 1955: 393). In other words, these controversies had little impact on mainstream thinking about the theory of the firm.

Another, later, challenge to the neoclassical model came from the managerial and behavioural theories of the firm. Fritz Machlup famously attempted to repel these attackers in his 1967 Presidential Address to the American Economics Association. He firstly argued that there was confusion as the role of the firm is price theory:

“[m]y charge that there is widespread confusion regarding the purposes of the “theory of the firm” as used in traditional price theory refers to this: The model of the firm in that theory is not, as so many writers believe, designed to serve to explain and predict the behavior of real firms; instead, it is designed to explain and predict changes in observed prices (quoted, paid, received) as effects of particular changes in conditions (wage rates, interest rates, import duties, excise taxes, technology, etc.). In this causal connection the firm is only a theoretical link, a mental construct helping to explain how one gets from the cause to the effect. This is altogether different from explaining the behavior of a firm. As the philosopher of science warns, we ought not to confuse the explanans with the explanandum”. (Machlup 1967: 9).

He then went on to argue that those behavioural and managerial theorists who were attacking the neoclassical model were doing so erroneously since they were working at a different level of analysis relative to that of the neoclassical model. The behavioural and managerial theories are aimed at the level of the individual firm whereas the neoclassical model concerns the industry level and thus, Machlup argued, the former are not genuine theoretical rivals to the latter.³⁴

While there was clearly dissatisfaction with the neoclassical model earlier it was not until the 1970s that this dissatisfaction reached the point where mainstream economists started to challenge the neoclassical model as the standard theory of the firm.³⁵ It was only then that the pioneering efforts of Knight (1921b) and Coase (1937) were recognised and developed. It was work by Oliver Williamson (see, for example, Williamson 1971, 1973, 1975), Alchian and Demsetz (1972) and Jensen and Meckling (1976) that drove the upswing in interest in the firm as an significant economic institution. But there have always been those outside of the mainstream who have criticised the neoclassical model. Here we very briefly consider the two non-mainstream groups of models noted above, the behavioural models of the firm and the

³⁴A related ‘level of analysis’ attack has been made on the ‘present’ theories of the firm as has been noted by Foss and Klein (2008: 429): “[...] the critics are protesting the application of concepts designed for analysis of *markets exchange* to the study of firm organization”. That is, concepts appropriate at the market level are not appropriate at the firm level.

³⁵As Aghion and Holden (2011: 181) note, “[u]ntil the 1970s, the dominant theory of the firm was the neoclassical theory: namely, there are economies of scale (or scope) which justify that production activities up to some efficient scale (or up to efficient variety) be concentrated within one firm rather than scattered across multiple producers”.

managerial models of the firm.³⁶ These were some of the first models to look inside the black box of the neoclassical firm.

Behavioural and managerial models³⁷

Behavioural Models

'Behavioural models' of the firm have been developed since the 1950s. In these models it is assumed that there is a separation between ownership and control. Behavioural theorists consider the consequences of conflict between self-interested groups within firms for the way in which firms make decisions on price, output etc. The emphasis in these models is on the internal relations of the firm with little attention being paid to the external relations between firms.

Some of the seminal work on the behavioural theories can be traced back to Simon (1955). The theory has subsequently been developed by Cyert and March with whose names it has been connected right up to today.³⁸

In behavioural theory the corporation has a multiplicity of different goals. Ultimately these goals are set by top management via a continual process of bargaining between the groups within the firm. An important point here is that the goals take the form of aspiration levels rather than strict maximisation constraints. Attainment of the aspiration level 'satisfices' the firm: the behavioural firm's behaviour is 'satisficing' in contrast to the maximising behaviour of the traditional firm. The firm seeks levels of profits, sales, rate of growth etc that are 'satisfactory', not those that are maxima. Satisficing is seen as rational behaviour given the limited information, time and computational skills of the firm's management. The behavioural theory redefines rationality, rationality is now that of 'bounded rationality'.

Cyert and March argue that there are two sources of uncertainty that a firm has to deal with. The first is uncertainty that arises from changes in market conditions, that is, from changes in tastes, products and methods of production. The second is uncertainty arising from the behaviour of competitors. According to the behavioural theory the first form of uncertainty is avoided, as much as it can be, by search activity, by spending on R&D and by concentrating on short-term planning. A difference between the traditional and behavioural theories is the importance given in the behavioural theory to the short-run, at the expense of the long-run. To avoid competitor-originated uncertainty, Cyert and March argue that firms operate within a 'negotiated environment', that is, firms act collusively with their competitors.

The instruments the behavioural firm uses in decision-making are the same as in the traditional theories. Both theories consider output, price and sales strategy as the major instruments.³⁹ The difference between the theories lies in the way firm choose the values of these instruments. In the neoclassical theory such values are selected so to maximise long-run profits. In the behavioural theory the choice is made so that the outcome is the 'satisficing' level of sales, profits, growth etc.

The behavioural theory also assumes that the firm learns from its experience. In the beginning a firm isn't a rational institution in the neoclassical sense of 'global' rationality. In the long run the firm may tend towards global rationality but in the short run there is an important adaptive process of learning. Firms make mistakes, there is trial and error from which the firm learns. In a sense the firm has memory and learns via its past experience.

An aspect of the firm neglected by the traditional theory is the allocation of resources within the firm and the decision-making process that leads to that allocation. In the neoclassical theory the firm reacts to its environment, the market, while the behavioural theory assumes that firms

³⁶Lee (1984) argues that there is a connection between the "reformist" behavioural and managerial models of the firm and the "marginalist controversy" of the 1940s and 50s. But the connection resulted in no real change since as Mongin (1998: 280) notes "[...] it would be a mistake to believe that these writers [the reformists] were representative of the majority of the economics profession".

³⁷A good textbook discussion of these models is given in Sections E and F of Koutsoyiannis (1979).

³⁸The major reference for the behavioural model of the firm is Cyert and March (1963).

³⁹Sales strategy here includes all activities of non-price competition, such as, advertising, salesmanship, service, quality etc.

have some discretion and do not take the constraints of the market as definite and impossible to change. The important point here is that the behavioural theory looks at the mechanisms for the allocation of resources within the firm, while the neoclassical theory examines the role of the market, or price, mechanism for the allocation of resources between the different sectors of the economy.

The concept of 'slack' is used by Cyert and March to refer to payments made to groups within organisation over and above that needed to keep that group in the organisation. Slack is therefore the same as 'economic rent' accruing to a factor of production in the traditional theory of the firm. What is significant about the behavioural school is their analysis of the stabilising role of 'slack' on the activities of the firm. Changes in slack payments in periods of good and bad business means that the firm can maintain its aspiration levels despite the changes to its environment.

Managerial Models

Another group of models, from outside the mainstream, which have been developed mainly since the 1960s in an effort to overcome some of the shortcomings of the neoclassical model are the managerial models of the firm. These models are also based on the idea that there is a difference between ownership and control of the firm. It is argued that the managers of the firm have taken control of the firm away from the owners. The common theme running through this literature is that the managers of the firm pursue non-profit objectives, generally subject to a performance constraint involving a profit related variable.

One of the earliest and most influential works in the managerial revolution was Berle and Merns (1932). It was Berle and Merns who famously argued that firms were becoming manager controlled rather than owner controlled, as had been the case in the past. De Scitovsky (1943) was a proto-managerial model of the firm. He modelled an entrepreneur whose utility depends on income and leisure and who faces an income/leisure trade-off given by the firm's profit function. The entrepreneur maximises his utility at a point involving more leisure, and less profit, than the profit maximising point. More recent work, explicitly developing the managerial approach, are Baumol (1962, 1967), Williamson (1964, 1970) and Marris (1964).

The standard theory of the firm can be interpreted as assuming that the managers of the firm act purely for the good of the owners. Owners can control what the managers do and thus the managers maximise profits. There are no principal-agent problems. Managerial models, on the other hand, start from the twin ideas that ownership and control are separated and that managers, just like other economic agents, act in ways that promote their own interests. But within these models maximising assumptions are still maintained. The obvious question this gives rise to is, What is maximised?

This question has been addressed by Baumol (1962) in which it is assumed that managers maximise sales subject to a profit constraint and by Baumol (1967) in which he develops a dynamic model in which the firm's objective is to maximise growth. Marris (1964) also assumes growth maximisation subject to a rate of return constraint. While it may seem likely that profit maximisation and growth maximisation will lead to behavioural differences between the two, work by Robert Solow (see Solow (1971)) argues that each type of firm would react in qualitatively similar ways to parameter changes such as changes in factor prices, excise taxes or a profit tax.

Williamson (1964, 1970) assumes a more general managerial utility function. His managerial discretion models let managers make a trade-off between "slack" and profits. In the static version, slack can be taken either as excessive administrative staff or as managerial emoluments (corporate personal consumption). In the dynamic-stochastic version of Williamson's model, slack comes as in the form of internal inefficiency, which has much in common with Leibenstein's (1966) notion of X-inefficiency. Williamson claims that behaviour in his discretionary models is qualitatively different from that under profit, sales or growth maximisation, although Rees (1974), for example, disputes aspects of this claim.

Summary

The behavioural and managerial theories can be seen as an early attempt to develop a theory of the firm at the level of the individual firm, a theory which, as Oliver Williamson has said of the Cyert and March (1963) book, was an attempt “pry open what had been a black box, thereupon to examine the business firm in more operationally engaging ways”. (Williamson 1996b: 150).⁴⁰ But the success of this attempt was limited. Williamson’s interaction with people such as Herbert Simon, Richard Cyert and James March while he was at Carnegie-Mellon University did play a role in the development of the transaction cost theory of the firm (Williamson 1996b) but outside of this the behavioural/managerial theories have had little effect on the mainstream economic theories of the firm.⁴¹

Demsetz and the neoclassical model

As noted above the neoclassical model held sway in mainstream economics up until 1970s and even today is still the one model of the ‘firm’ that every economist knows. In fact it’s likely to be the only model of the firm they do know. The standard interpretation of the neoclassical model is one in which firms, in the Coaseian sense, do not exist. The model is one of zero transaction costs in which agents interact with each other only via the price mechanism and elaborate (complete) contracts. Harold Demsetz is one author who disagrees with this interpretation of the neoclassical model. For him the ‘firm’ in the neoclassical model is a specialised production unit, specialised in the sense that it produces only for those outside the firm.

Demsetz (1995: First commentary) argues that the neo-classical model offers both a definition of the firm and a rationalisation for the existence of firms, but he admits that these are mostly implicit. Demsetz starts by noting that the problem that the neoclassical model tackles is to see how the price system works and how it is able to deal with the interdependencies of the modern economy. The theory sets out to do this by envisioning a hypothetical economy within which people must depend on others. Demsetz (1995: 7) explains,

“[t]he construction depends on two characteristics of economic activity: extreme decentralization and extreme interdependency. Extreme decentralization deprives all firms and households of influence over price. So they do not set price; the system does. This aspect of neoclassical theory is well understood. The need for interdependency is not”.

The opposite of interdependency is self-sufficiency, by which Demsetz means production for one’s own consumption. Robinson Crusoe stranded alone on an island must be self-sufficient; there is no one else to depend on. The neoclassical economy is one in which there is no self-sufficiency so that all people in this hypothetical economy are dependent on all other people in the economy. That is, there is extreme interdependency. Demsetz argues that

“[t]his is accomplished with the aid of two “black boxes”: the household and the firm. The household sells its services to others and buys goods from others. It does not self-employ resources to produce goods for its own members; it offers its resources to firms. Firms buy or rent these resources, and they produce goods that are not for consumption by their owners and employees as such, but are for exclusive sale to households. The role of prices in accommodating this high degree of interdependency is of interest, not the manner in which households and firms manage their internal affairs. The contribution made by the household and the firm in this theory is to

⁴⁰For retrospective look at *A Behavioral Theory of the Firm* after 45 years see Augier and March (2008).

⁴¹If you look at the standard microeconomics textbooks, both undergraduate and graduate, it is difficult to find a discussion of either behavioural or managerial models. Koutsoyiannis (1979) is one of the few that gives serious attention to these models, and it is now more than 30 years old. The impact of these works may have been greater in management than economics. Argote and Greve (2007: 337), for example, claim that *A Behavioral Theory of the Firm* “continues to be one of the most influential management books of all time”.

make the price system deal with extreme interdependency and decentralization. “In-the-household” production and “on-the-job” consumption are ruled out”. (Demsetz 1995: 8).⁴²

The production unit in the neoclassical economy is specialised in the sense that it produces for those outside the firm, so that the firm is not just a black box, it is a specialised black box. There is no discussion of the managing of production. The role of the firm in the neoclassical theory is to separate production from consumption so that there is no self-sufficiency. The coordination of production and consumption is achieved via two factors: first, impersonally determined market prices and second, personally defined tastes. The neoclassical model lays out the nature of the interactions between these components. Thus the perfectly competitive firm is one important ingredient in a scenario in which the price system is the only coordination mechanism for harmonising production and consumption.

Demsetz goes on to note that the internal organisation of the firm is not addressed in the neoclassical theory. The firm need not be an organisation at all, a single owner/manager/employee is all that is required. For Demsetz the neoclassical firm is no more or less than a specialised unit of production. The important criterion for the neoclassical firm is that it separates production from consumption with production being exclusively for consumption by those outside the firm.

In the neoclassical world in which everyone possesses perfect information about prices and technologies, each owner of resources can manage their own resources, placing them in their highest value uses in response to the prices that they face. These resource owners can write any contracts needed to coordinate their relationships.

Demsetz then makes the point that this view of the firm is very different from that of either Knight or Coase or from the modern theory of the firm literature, which follows, in the main, from Coase. In the Coaseian literature markets and firms are seen as substitutes, in that as transaction costs fall the market is used more and firms do less. In the limit as transaction costs go to zero the firm ceases to exist and all activities take place via markets. In the Demsetz framework the relationship between firms and markets is complementary. As transaction costs fall, the costs to specialisation fall as the use of the market becomes cheaper and more specialisation takes place and thus more firms are created. As transaction costs increase, the use of the market becomes more expensive and thus it is used less, self-sufficiency become more common and the number of firms falls.

Demsetz sums up the specialisation theory of the firm as,

“[t]he bottom line of specialization theory is that *firms exist because producing for others, as compared to self-sufficiency, is efficient; this efficiency is due to economies of scale, to specialized activity, and to the prevalence of low, not high, transaction costs*”. (Demsetz 1995: 11; emphasis in the original).

One interesting implication of the specialisation theory is that it guarantees profit maximisation. Given that firms only produce for sale to those outside the firm, there can be no on-the-job consumption and thus the owner of the firm maximises utility by maximising profits. As there can be no utility gained from on-the-job consumption the owner maximises utility by having the firm maximise profit and then saving or consuming this profit in his role as consumer.

⁴²This separation between the household and firm is also noted by Hicks (1946: 79): “[...] the enterprise (the conversion of factors into products) may be regarded as a separate economic unit, detached from the private account of the entrepreneur. It acquires factors, and sells products; its aim is to maximize the difference between their value”. Spulber (2009: 125) calls this separation of the firm’s objectives and the consumer’s objectives the “neoclassical separation theorem”, which he says makes three assertions: “(1) firms maximise profits, (2) firms generate gains from trade compared to autarky, and (3) firm decisions are separate from consumer decisions”. For expanded discussion see Spulber (2009: 127-32). For Spulber the firm “is defined to be a transaction institution whose objectives differ from those of its owners”. (Spulber 2009: 63). The importance of this separation is noted by Mas-Colell et al (1995: 153) when they observe “[i]f prices may depend on the production of the firm, the objective of the owners may depend on their tastes as consumers”. This implies that the objective of profits maximisation by the firm may be lost.

Summary

For the most part the classical economists utilised a theory of aggregate production, and distribution, not a theory of firm level production or a theory of the firm. The neoclassical economists sort to develop a theory of firm level production, but it can be seen as one without firms. The 'present' theories of the firm are an attempt to create a theory of the firm. To lay a foundation for our review of the current approaches we first survey the founding works on which they are largely based: Knight (1921b) and Coase (1937).

The founding works

Knight 1921(b)

Demsetz (1988b: 244) goes so far as to state “[...] it can be said without hesitation that Knight launched the modern theory of the firm in 1921”. However the primary motivation of Knight (1921b) wasn't to examine the organisation of the firm or explain the existence of the firm, it was to explain the existence of profit. The theory of the firm was a byproduct of his explanation of profit. Although as Foss (2000: xix) notes “[...] the connection between his theory of profits and his theory of the firm is not entirely clear”.

The standard view of Knight's rationale for the existence of the firm, see for example Demsetz (1995: 2-4), doesn't depend on profit, but on risk, or more accurately, risk redistribution. The entrepreneur forms a firm as a way of specialising in risk-taking. Employees receive a stipulated income and the entrepreneur takes the residual income of the firm and thereby bears most of the risk associated with uncertainty about the future. The advantage of the firm, according to the standard view, is that there are gains to be made from this redistribution of risk between the entrepreneur and the firm's employees. The profit and loss consequences of fluctuations in the business outcomes can be better absorbed by the entrepreneur than the employees. The entrepreneur contracts to pay a fixed wage to workers, thereby protecting them from the fluctuations in business outcomes. Knight sees this as efficient since the entrepreneur is less averse to bearing risk. Presumably, risk is not handled as well without firms.

Another view is offered by Boudreaux and Holcombe (1989).⁴³ They see Knight's theory of the firm as stemming from the role of the entrepreneur as the person who decides what to produce or whether or not to introduce a new production process in a world of Knightian uncertainty. For Knight, the goods and services to be produced are not given, as in the neoclassical theory, thus entrepreneurs must make a decision as to which goods to produce. Given that the entrepreneurs face a world of uncertainty, such decisions must be made on the basis of 'intuitive judgement'. The need for 'judgement' is due to the entrepreneur having to deal with uncertainty resulting from the fact that prices of the outputs are unknown when the decisions about production are made. This price uncertainty is the result of changing consumer desires and the uncertainty as to the reactions of competitors. Entrepreneurs differ from non-entrepreneurs in that entrepreneurs receive the return from 'judgement', that is, entrepreneurs receive the residual (positive or negative) left after the costs incurred at the time the production decision was made are subtracted from revenues.

For Boudreaux and Holcombe the “distinguishing characteristic of the Knightian entrepreneur [...] is that he makes decisions under uncertainty about how resources will be allocated”. (Boudreaux and Holcombe 1989: 152). The Knightian firm's primary function is, in Boudreaux and Holcombe's view, entrepreneurial, decisions must be made without the guidance of market prices since the market doesn't exist yet. Entrepreneurial activity is necessary for the development of markets. New goods create new markets. For Knight, the products to be produced is a decision made within the firm. The entrepreneur is the person in the firm who makes such decisions. Thus for Boudreaux and Holcombe the Knightian theory of the firm is driven by a theory of the entrepreneur, this they claim differentiates the Knightian theory from that of Coase, who they

⁴³See Foss (1993) for criticism of Boudreaux and Holcombe.

argue puts forward a theory of management that leaves no room for genuine entrepreneurship. For Boudreaux and Holcombe, the Knightian firm exists in order to facilitate decision making in a world of true uncertainty, that is, to facilitate true entrepreneurial decision making.⁴⁴ Presumably, such decision making is not as efficient without firms.

Barzel (1987) and McManus (1975) put forward a moral hazard explanation for the Knightian firm. The firm arises here “because, for certain kinds of risks, the functions of risk taking and management are inseparable due to the prohibitively high costs of enforcing constraints that would induce one individual, the manager, to maximize the wealth of another, the risk-taker”. (McManus 1975: 348). As noted in the redistribution of risk story above, firms are one way of specialising in risk-taking. Knight was aware of contractual and insurance arrangements as alternatives to the firm as ways of specialising in risk-taking but thought, because of the moral hazard problems, they were particularly costly to enforce in the case of risks of enterprise and hence the need for the creation of a firm. Presumably monitoring the manager is easier for the risk-taker in a firm than it is on the market.

An alternative view is given by Langlois and Cosgel (1993). Here it is argued that Knight’s theory of organisation has things in common with the more recent incomplete contracts approach to the firm.⁴⁵ Langlois and Cosgel summarise their view of Knight’s theory of organisation as

“[b]ecause of the non-mechanical nature of economic life, novel possibilities are always emerging, and these cannot be easily categorized in an intersubjective way as repeatable instances. To deal with this “uncertainty,” one must rely on judgment. Such judgment will be one of the skills in which people specialize, yielding the usual Smithian economies. Moreover, some will specialize in the judgment of other people’s judgment. As the literature since Coase [1937] suggests, however, a theory of specialization is not by itself a theory of organization, since, in the absence of transaction costs, there is no reason why the division of labor could not be undertaken through markets rather than within a firm. Knight’s answer is that the function of judgment is ultimately non-contractible”. (Langlois and Cosgel 1993: 462).

The non-contractibility of judgement leads to the entrepreneur’s skills not being tradable on markets, thus the division of labour cannot “be undertaken through markets rather than within a firm” and hence the need for the firm.⁴⁶ The optimal organisational structure that results from this has the entrepreneur being the residual claimant, and he hires the other agents for a fixed payment. Langlois and Cosgel argue that incompleteness results in the entrepreneur owning the other assets in the firm on the assumption that the entrepreneur’s participation is the most important to the resulting joint product. If we compare this case with that of the standard risk redistribution case noted above, we see that the residual claimant doesn’t so much insure the other agents, as in the risk redistribution story, rather it is simply that, due to the non-contractibility, the optimal arrangement is for the entrepreneur to receive the residual and the other agent to receive a fixed payment.

Thus if one wished to write Whig History, Knight’s theory of the firm would be a forerunner, not of the theory of moral hazard and asymmetric information, but of the incomplete contracts approach to vertical integration. Langlois and Cosgel contend that Knight saw the causes of incompleteness in the lack of knowledge of the categories of action and the consequent need for judgement. For Knight incompleteness of contract was ultimately a matter of uncertainty.

⁴⁴Foss (1993: 273) conceptualises this as “the firm and vertical integration exist because entrepreneurs cannot communicate—without exorbitant information costs—their idiosyncratic ‘versions’ (innovations) to owners of assets necessary for realizing this vision; therefore, they integrate such activities”. This inability to communicate with assets owners means that it is difficult to hire assets on the market and thus the need for entrepreneurs to supply the needed assets themselves by forming a firm.

⁴⁵See Grossman and Hart (1986, 1987), Hart and Moore (1990) and Hart (1995).

⁴⁶Foss and Foss (2006: 2) note “[...] there is no market for judgment that entrepreneurs rely on, and therefore exercising judgment requires the person with judgment to start a firm”.

Coase 1937

Coase opens the paper by pointing out that there has been, in economics, a failure to clearly state the assumptions on which theories are built. He notes that two questions can be asked of a set of assumptions: Are they tractable? and Do they correspond with the real world? Coase argues that it is important to have a clear definition of the word “firm” since much economic analysis starts with the individual firm rather than the industry and it is important to know the difference between the theoretic firm and the real world firm. The aim of Coase’s paper was to provide a definition of the firm that “is not only realistic in that it corresponds to what is meant by a firm in the real world, but is tractable by two of the most powerful instruments of economic analysis developed by Marshall, the idea of the margin and that of substitution, together giving the idea of substitution at the margin”. (Coase 1937: 386-7).

Coase begins his search for a definition of the firm by pointing out that the standard treatment of the economic system is one where the price mechanism provides all the coordination of resources required. Resource allocation is dependent directly on the price mechanism. But in the firm, Coase notes, the price system does not allocate resources, authority does. The use of authority to supersede the price mechanism is, in Coase’s view, the distinguishing mark of the firm. Coase then asks, If all coordination can be done by the price mechanism, why is the firm, with its coordination by authority, necessary?

In Section II, Coase states the the task ahead is to explain why a firm would emerge in a specialised exchange economy. He first points out that it could emerge if it was desired for its own sake. It could arise if some people preferred working under the direction of others or if some people wished to control others. Also firms may arise if customers preferred goods produced in this way to goods produced by other institutional arrangements. However, Coase points out that these motivations can not explain all firms we see, hence there must be other factors involved.

The, now, most famous “other factor” is that there are costs to using the price mechanism. To quote Coase (1937: 390-2):

“[t]he most obvious cost of “organising” production through the price mechanism is that of discovering what the relevant prices are. [. . .] The costs of negotiating and concluding a separate contract for each exchange transaction which takes place on a market must also be taken into account. [. . .] It is true that contracts are not eliminated when there is a firm but they are greatly reduced. [. . .] There are, however, other disadvantages-or costs-of using the price mechanism. It may be desired to make a long-term contract for the supply of some article or service. [. . .] Now, owing to the difficulty of forecasting, the longer the period of the contract is for the supply of the commodity or service, the less possible, and indeed, the less desirable it is for the person purchasing to specify what the other contracting party is expected to do. [. . .] When the direction of resources (within the limits of the contract) becomes dependent on the buyer in this way, that relationship which I term a “firm” may be obtained. A firm is likely therefore to emerge in those cases where a very short term contract would be unsatisfactory. [. . .] We may sum up this section of the argument by saying that the operation of a market costs something and by forming an organisation and allowing some authority (an “entrepreneur”) to direct the resources, certain marketing costs are saved. The entrepreneur has to carry out his function at less cost, taking into account the fact that he may get factors of production at a lower price than the market transactions which he supersedes, because it is always possible to revert to the open market if he fails to do this”.

Coase also notes that the different treatment of in house and market transactions by Government and regulatory bodies could also explain why some firms exist. Having explained why a firm could exist, Coase goes on to note that a firm consists of the relationships that are brought into existence when the control of resources is dependent on an entrepreneur.

An advantage of the approach just developed, claims Coase, is that it is possible to give a meaning to a firm becoming larger or smaller. A firm becomes larger when a transaction that

could be carried out in the market, is instead organised by the entrepreneur. The firm, therefore, becomes smaller when the entrepreneur gives up organising such a transaction.

Next Coase considers the question as to why, if by creating a firm, the costs of production can be reduced, are there any market transactions at all. “Why is not all production carried on by one big firm?” (Coase 1937: 394). The answer according to Coase is, first, that as a firm gets bigger there may be decreasing returns to entrepreneurial activity. That is, the cost of an additional transaction being organised within the firm may rise. Secondly, as the number of transactions which are organised in house increases, the entrepreneur may fail to place the factors of production in the uses where their value is maximised. This means the entrepreneur fails to make the best use of the available factors of production. Finally the supply price of one or more of inputs to production may increase, because the “other advantages” of a small firm are greater than those of a large firm.⁴⁷ As a result of these factors, “a firm will tend to expand until the costs of organising an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organising in another firm”. (Coase 1937: 395).

Coase (1937: 396-7) then notes that all other things being equal, a firm will tend to be larger:

- (a) the less the costs of organising and the slower these costs rise with an increase in the transactions organised.
- (b) the less likely the entrepreneur is to make mistakes and the smaller the increase in mistakes with an increase in the transactions organised.
- (c) the greater the lowering (or the less the rise) in the supply price of factors of production to firms of larger size.

An additional reason for why efficiency will decrease as the firm grows larger is that as more transactions are controlled by an entrepreneur, these transactions are likely to be either different in kind or different in place. Mistakes in decision making are more likely as there is an increase in the spatial distribution of transactions, the dissimilarity of the transaction and in the probability of changes in prices relevant to production. Changes which lessen the spatial distribution between transactions will lead to an increase in the size of the firm, as will improvements in managerial technique.

The ideas of “combination” and “integration” can be given precise meaning using the analysis presented above. Combination is when transactions normally undertaken by two or more entrepreneurs are undertaken by one and this turns into integration when the transaction was previously carried out on the market. Firms can grow via either or both of these two ways.

In the last section of the paper Coase asks whether the concept of the firm he has developed is realistic and manageable? As to realism, he contends the best way to see what constitutes a firm in practice is to look at the legal relationship between the master and servant or employer and employee.⁴⁸ The essentials of the employer and employee relationship is given by Coase as follows:

1. The servant must be under the duty of rendering personal services to the master or to others on behalf of the master, otherwise the contract is a contract for sale of goods or the like.

⁴⁷This point is explained by Coase in footnote 1 on page 395, which reads: “[f]or a discussion of the variation of the supply price of factors of production to firms of varying size, see E. A. G. Robinson, *The Structure of Competitive Industry*. It is sometimes said that the supply price of organising ability increases as the size of the firm increases because men prefer to be the heads of small independent businesses rather than the heads of departments in a large business. See Jones, *The Trust Problem*, p. 531, and Macgregor, *Industrial Combination*, p. 63. This is a common argument of those who advocate Rationalisation. It is said that larger units would be more efficient, but owing to the individualistic spirit of the smaller entrepreneurs, they prefer to remain independent, apparently in spite of the higher income which their increased efficiency under Rationalisation makes possible”.

⁴⁸In a footnote Coase explains that the legal concept of employer and employee and the economic concept of a firm are not identical. He notes that the firm may imply control over another person’s property in addition to their labour. But the identity of these two concepts is sufficiently close for an examination of the legal concept to be of value in appraising the worth of the economic concept. However in Coase (1988a: 37) he writes “I consider that one of the main weaknesses of my article stems from the use of the employer-employee relationship as the archetype of the firm. It gives an incomplete picture of the nature of the firm. But more important, I believe it misdirects our attention”.

2. The master must have the right to control the servant's work, either personally or by another servant or agent. It is this right of control or interference, of being entitled to tell the servant when to work (within the hours of service) and when not to work, and what work to do and how to do it (within the terms of such service) which is the dominant characteristic in this relation and marks off the servant from an independent contractor, or from one employed merely to give to his employer the fruits of his labour. In the latter case, the contractor or performer is not under the employer's control in doing the work or effecting the service; he has to shape and manage his work so as to give the result he has contracted to effect. (Coase 1937: 403-4).

It is noted by Coase that what distinguishes an agent from an employee is not the presence or absence of a fixed wage or the payment only of commission, but rather the freedom with which an agent may carry out his employment. Coase argues that it is the fact of direction that is the essence of the legal concept of the employer and employee relationship just as it was in the economic concept of the firm he developed. He concludes that his definition is therefore realistic. The question is then asked, Is it manageable? Again the answer is yes, the principle of marginalism works smoothly. The question is, Does it pay to organise an additional transaction under a given entrepreneur? Should the transaction be undertaken by this firm or some other firm or in the market. At the margin the cost of undertaking the transaction in any given firm will equal the cost to either another firm or in the market.

Summary

It is normally the work of Knight and Coase that is credited with providing the foundations for the development, starting around 1970, of the current theories of the firm. Knight's work can be seen as, depending on interpretation, arguing that a firm is needed to facilitate risk redistribution or because of the need for intuitive judgement in a world of uncertainty or as a catalyst for moral hazard and asymmetric information theory or as a catalyst for incomplete contract theory. So while many authors argue for Knight as a founder of the modern approaches to the firm, they can not agreed on why. The importance of Coase (1937) stems from the fact that it was the starting point for much of the contractual literature on the firm, and the theory of economic organisations more generally, insomuch as it is in this paper that we see the main questions underlying the modern theory of the firm being raised together for the first time. Coase sets out to "discover why a firm emerges at all in a specialized exchange" – a question about the existence of the firm; he also sets out to "study the forces which determine the size of the firm" – an issue to do with the boundaries of the firm; and he inquires into the reasons for "diminishing returns to management" – issues to do with the internal organisation of the firm. It was the efforts to answer these questions that initiated the charge from seeing the theory of the firm as just part of price theory to seeing it as an important topic in its own right. Coase also provides one of the main building block for answers to these issues, the "costs of using the price mechanism" or transaction costs.

But as Coase notes "[t]he article was not an instant success". (Coase 1988a: 23). In fact it took nearly 40 years for it to become an overnight success. Before 1970 the paper was, in Coase's own words, "an article much cited and little used". (Coase 1972: 63). Coase argues that this changed both with the publication of "The Problem of Social Cost" (Coase 1960) which helped rekindle interest in 'The Nature of the Firm' via the greater appreciation of transaction costs it brought about and with the writings of Oliver Williamson who incorporated transaction costs into the analysis of the distinction between hierarchy and markets. (Coase 1988b: 34-5).

It is to this post-1970 literature, developed by Williamson and others, we now turn.

The ‘present’

The post-1970 theories of the firm⁴⁹

Foss, Lando and Thomsen (2000: 634) offer an useful classification, which we follow here, of the mainstream post-1970 economics literature on the theory of the firm. They partition the theory into two general groups:

1. Principal-agent type models where agents can write comprehensive contracts characterised by ex ante incentive alignment under the constraints imposed by the presence of asymmetric information.
2. Incomplete contracts models which are based on the idea that it is costly to write contracts and thus contracts will have holes or inefficient provisions, and therefore there is a need for ex post governance.⁵⁰

This division can be seen as resulting from the breaking of two different assumptions embedded in the general equilibrium (Arrow-Debreu) version of the neoclassical model.⁵¹ The first group corresponds to the breaking of the assumption that there are no asymmetries of information between parties and thus no principal-agent problems, of either the adverse selection or moral hazard kind. The second grouping results from breaking the assumption that agents can foresee all future contingencies and can costlessly contract on all such eventualities. We discuss each group in turn.

Within the principal-agent classification Foss, Lando and Thomsen (2000: 636-8) identify three sub-groups: 1) the nexus of contracts view, 2) the firm as a solution to moral hazard in teams approach and 3) the firms as an incentive system view.

The nexus of contracts view was developed in papers by Alchian and Demsetz (1972), Jensen and Meckling (1976), Barzel (1997), Fama (1980) and Cheung (1983). The important innovation here was the recognition that it is difficult to draw a line between firms and markets, firms are seen as a special type of market contracting. What distinguishes firms from other forms of market contract is the continuity of the relationship between input owners.

Most famously in the Alchian and Demsetz version of this approach, they argue that the authority relationship between the employer and employee is in no way the defining characteristic of a firm.⁵² The employer has no more authority over an employee than a customer has over his grocer. “Firing”, of either the employee or grocer, is the ultimate punishment that either the employer or customer can use in cases of “disobedience”. Alchian and Demsetz argue that, in economic terms, the customer “firing” his grocer is no different from the employer firing his employee. In both cases one party stops dealing with the other, terminating the “contract” between them. In this approach the firm is seen as little more than a nexus of contracts, special in its legal standing and characterised by long term nature of the relationship between the input owners. In this approach it is not generally useful to talk about firms as distinctive entities, a nexus of contracts could be called more firm-like if, for example, the residual claimants belong to a concentrated group but the term “firm” has little meaning beyond this.

Roberts (2004: 104) responds to this line of argument:

⁴⁹This section draws on material from Walker (forthcoming).

⁵⁰More precisely, contracts can be incomplete for two reasons; the first is that the contract is “insufficiently state contingent” in that its terms are not optimal in all possible states of the world, while the second is where the contract is “obligationally incomplete” in that it has a gap or missing provision. For example, a contract may state that a supplier must provide one widget in all states of the world rather than the optimal number of widgets, which varies with the state of the world; or it may not specify what is to happen if a supplier’s factory is destroyed in an earthquake.

⁵¹The Arrow-Debreu framework was not originally conceived as a theory of contracting per se, but rather it was seen as an analytical apparatus for modelling competitive equilibrium. But the efficiency properties associated with trade involving complete contingent claims contracts - that is, contracts specifying the price, date, location and physical characteristics of a commodity for every future state of nature - made such contracts the standard against which other, more realistic, contracts are compared.

⁵²In later work Alchian has stated that this assertion is incorrect (Alchian 1984: 38) while Demsetz (1995: 37) claims the idea “is a mere aside” in their paper.

“[w]hile there are several objections to this argument, we focus on one. It is that, when a customer “fires” a butcher, the butcher keeps the inventory, tools, shop, and other customers she had previously. When an employee leaves a firm, in contrast, she is typically denied access to the firm’s resources. The employee cannot conduct business using the firm’s name; she cannot use its machines or patents; and she probably has limited access to the people and networks in the firm, certainly for commercial purposes and perhaps even socially”.

The second grouping, the “firm as a solution to moral hazard in teams approach”, was developed by Alchian and Demsetz (1972) and Holmström (1982). Alchian and Demsetz (1972) extend their discussion, outlined above, by noting that the firm is more than just a special legal arrangement, it is also characterised by team production. The problem that arises here is that with team production, the marginal products of the individual members of the team are hard to measure. This means that free-rider behaviour is now possible since team production can act as a cover for shirking. The Alchian and Demsetz solution is to give the right to hire and fire the members of the team to a monitor who observes the employees and their marginal products. To ensure that the efficient amount of monitoring takes place, the monitor is given the rights to the residual income of the team.

Holmström (1982) looks at the incentive problems to do with monitoring and identifies possible solutions.⁵³ Holmström assumes that the members of the team each take actions which are unobservable to the monitor but the overall result of the combined actions is observable. What Holmström shows is that it is only under very restrictive assumptions that the monitor can ensure that efficient effort levels will be provided by each team member. The way the monitor would ensure this is to design a sophisticated incentive scheme. But Holmström shows that given unobservable effort levels, the requirements of the incentive scheme being a Nash equilibrium, budget balancing and Pareto optimality, can not be met.⁵⁴ More specifically, a budget-balancing incentive scheme can not reconcile Nash equilibrium and Pareto optimality. This is because each team member will equalise the costs and benefits of extra effort: that is, if the team revenue is increased by the efforts of a single member, that member should receive that revenue to ensure that they are properly motivated. But as the monitor only knows that team revenue has increased and not the effort levels of each individual member, all members of the team would have to each receive the extra revenue to ensure that the hard working member is rewarded for his efforts. But this will, obviously, violate the balanced budget condition. This suggests that there is an advantage, in terms of incentives, in the team not having to balance their budget.⁵⁵

Clearly the role of the “monitor” in the Alchian and Demsetz model is very different to their role in the Holmström model. For Alchian and Demsetz, the monitor oversees the behaviour of the team members, in the Holmström model, the monitor injects the capital needed so that the team members do not have to balance their budget.

The third subgroup is the “firms as an incentive system view”. Early contributors to this approach were Holmström and Milgrom (1991, 1994). In Holmström and Milgrom (1994) it is stressed that the firm should be viewed as ‘a system’, that it is a set of contractual relationships which endeavour to mitigate incentive problems. In their view the firm is characterised by a number of factors: 1) the employees do not own the non-human assets of the firm; 2) the

⁵³Importantly Holmström ignores team synergies by assuming an additive production function.

⁵⁴Budget balancing means that the incentive scheme has to fully distribute the revenues among the team members.

⁵⁵McAfee and McMillan (1991: 562) describe the problem and one solution mechanism as,

“[t]he principal offers to pay each of the n agents 100 percent of any marginal increase in team output. Clearly this gives each agent the appropriate incentive to exert effort. It does, however, result in the principals total variable payment being n times the value of output. To balance this, the fixed part of the payment function must be negative: in fact, in this case each agents fixed payment is set equal to the expected value of output minus the agents production cost, so that the agents earn zero rents on average. Thus the optimal contract has the principal initially (before production takes place) collecting money from the agents, and then (after the production process) paying each agent the full value of the teams output. In other words, in the presence of moral hazard, the principal achieves his ideal outcome by, at the margin, breaking the budget: by eliminating the requirement that the marginal payments sum to one, and by manipulating the lump-sum payments instead of the marginal payments [. . .]”.

employees are subject to a 'low-powered incentive scheme' (see below); and 3) the employer has authority over the employee.

Holmström and Milgrom (1991) make two observations. First, they note that there are a number of ways that an employee can spend their time, many of which can be of value to an employer. But if these multiple activities compete for the worker's attention then the incentives offered for each of the activities must be comparable. Otherwise, the employee will put most effort into those things that are most well compensated and put less effort into the others activities. The second observation relates to the provision of strong incentives to a risk-averse employee. Providing strong financial incentives is costly because it loads extra risk into the worker's pay. In addition, the cost is greater the more difficult it is to measure performance. This means that, other things being equal, tasks where performance is hard to measure should not be given as intense incentives as ones that are more accurately observed. But having low-powered incentives means that the employer needs to be able to exercise authority over the use of the employee's time, since the employee will not have the proper incentives to be productive.

This logic suggests that, conversely, an independent contractor should face the opposite combination of instruments. The choice between having an employee or using an independent contractor depends on the ability of the principal to measure each dimension of the agent's contribution. Thus, in the Holmström and Milgrom approach, measurability of performance is one important determinant of the boundaries of the firm. In addition their approach incorporates the importance of the allocation of property rights to the physical assets in determining incentives via determination of bargaining positions as is the case with the Williamson and Grossman-Hart-Moore approaches, both of which are discussed below.

In the incomplete contracting theories group Foss, Lando and Thomsen (2000: 638-43) identify five subgroups: 1) the authority view, 2) the firm as a governance mechanism, 3) the firm as an ownership unit, 4) implicit contracts and 5) the firm as a communication-hierarchy. We will add a sixth group, 'the reference point approach'.

In the authority view, the firm is seen as being defined as an employment relation. This view is one example were an approach to the firm founded in the pre-1970 period, this approach is most closely associated with Coase (1937) and Simon (1951), is still developed in the current mainstream. As noted in more detail in Section , for Coase a firm will arise when it is cheaper to carry out a transaction in a firm than it is to do so over the market. Given it costs something to enter into a market contract, that is, there are transaction costs, firms will emerge to carry out what would otherwise be a market transaction when it is cheaper for the firm to handle that transaction. The size of the firm (the boundaries of the firm) will be determined when the cost of organising a transaction within the firm equals the cost of using the market. Coase notes that within the firm contracts are not eliminated but are greatly reduced and the nature of the contract changes. When a factor of production is employed within the firm the contract controlling it is incomplete. The factor (or its owner) agrees, for remuneration, to obey the directions of the manager of the firm, within certain limits. In the last section of Coase (1937), it is noted that the relationship that constitute the firm corresponds closely to the legal concept of the relationship between the employer and employee. Coase explains that "direction" is the essence of the legal concept of the employment relationship, just as it is for the concept of the firm that he developed.

For Simon (1951) the issue is a comparison of an employment contract against a contract between two autonomous agents. A contract between autonomous agents specifies an action to be taken in the future along with its price while an employment contract specifies a set of acceptable instructions that the employee has to accept if asked to carry them out by the employer. The advantage of the employment contract is its flexibility, the employer does not have to pre-commit to an action and can adapt the choice of action to the state of the world that occurs.

A more modern approach to these issues is Wernerfelt (1997). He compares three alternative governance structures (game forms) for situations where a buyer needs a sequence of human asset services: 1) the hierarchy game form, 2) the price list game form and 3) the negotiation-as-needed game form.⁵⁶ An employee is defined as someone who sells his services in a specific game

⁵⁶Wernerfelt (1997: 490) introduces the game forms with three simple examples:

1. As a typical day unfolds, you learn that you will need several services from your secretary. In principle,

form characterised by the absence of bargaining over adaptations to changing circumstances. The firm is seen as consisting of the buyer of human asset services, along with a set of sellers, provided that the human services are traded in the “employment relationship” or “hierarchy” game form. The hierarchy game form is defined as the situation in which the parties engage in once-and-for-all wage negotiation, the manager describes desired services sequentially, and either party may terminate the relationship at will. In this model, the boundaries of the firm are given by the set of agents employed by the buyer. Whether one uses the employment relationship or an alternative game form depends on the nature of the expected adaptations. If many diverse and frequent adjustments are needed, the employment relationship involves lower adjustment costs than any of the other governance structures. The price list game form is better when the list of possible adjustments is small and the negotiation-as-needed game form is better when adjustments are needed infrequently.

The ‘firm as a governance mechanism’ approach is most commonly associated with the work of Oliver Williamson (see, for example, Williamson 1971, 1973, 1975, 1979, 1985, 1996a). Williamson’s work is based on the twin notions of bounded rationality, which results in contractual incompleteness, and opportunism, thought of as ‘self-interest with guile’. An upshot of these ideas is that contractual agreements need various kinds of safeguards built into them. For example, contractual agreements could involve ‘hostages’, that is, one party may post a bond with the other. The contractual arrangements and their associated safeguards are referred to as ‘governance structures’ by Williamson. This basic idea is that transactions can be assigned to governance structures on the basis of their transaction properties.

Gibbons (2005: Sections 1.1, 1.4, 2.2, 3) argues that Williamson’s works can be read as suggesting two elemental theories of the firm – rent-seeking and adaptation. In the rent-seeking theory of the firm integration can be the efficient governance structure because integration can put a stop socially destructive haggling over ‘appropriable quasi-rents’ (or AQRs). Williamson argues that within firms conflicts are settled by fiat, which can be a more efficient way to handle such issues than haggling. The basic logic is that, in the presence of AQRs, non-integration cannot avoid inefficient haggling because even though the haggling is jointly and socially unproductive, it is a source of private pecuniary gain, and thus integration, which brings with it dispute-resolution by fiat – and thus less haggling – can be more efficient. Note also the result that the larger are the AQRs the more likely is integration, presumably because the socially unproductive haggling is either more likely or more costly (or both) the larger are the AQRs.⁵⁷

In Gibbon’s adaptation version of Williamson’s theory of the firm, integration can be the optimal governance structure for transactions that require adaptive, sequential decision-making in situations where uncertainty is resolved over time. To formulate an adaptation theory, a setting – in which uncertainty is endemic – must be created such that neither *ex ante* contracts nor *ex post* renegotiation can achieve post-uncertainty first-best adaptation. A second-best solution may be to concentrate authority in the hands of a ‘boss’ who makes decisions – albeit self-interested decisions – after the resolution of any uncertainty. Thus for the adaptation (and the rent-seeking) theory the emphasis is on authority and control, where the ‘boss’ makes any necessary decisions. This is in contrast to the ‘firm as a solution to moral hazard in teams approach’

the two of you could contract over the provision of each service as its nature becomes clear. However, under such an arrangement you would spend a lot of time negotiating. We therefore have the institution normally called the employment relationship under which the secretary has agreed to supply *ex ante* unspecified services for a certain number of hours.

2. Consider what happens if a general contractor remodels your house. You may change your mind during construction, but because these adaptations are infrequent they are typically handled through negotiation on an *as-needed* basis.

3. Suppose next that you are at H & R Block getting help with your tax return. While at the store, you may realize that you need additional services: there may be more schedules to file or you may want to prepay part of next year’s taxes. In this case you know the price of each adaptation *ex ante* and no new negotiation is needed. Since the number of possible adjustments is small, the price list governs adaptation cheaply.

⁵⁷Strictly speaking this can’t be the whole story since while larger AQRs may increase the costs of non-integration, until we know something about the costs of integration, little can be said about the trade-off between the two and the probability of integration.

which concentrates on incentives and thereby ignores control and the ‘firm as an ownership unit’ approach which blends the two but it does mean that the adaptation theory shares similarities with ‘the authority view’ discussed above.⁵⁸ Williamson argues that it is only when there is a need to make unforeseen adaptations does the market versus internal organisation question become an interesting issue.

Within the rent seeking framework it is often argued that AQRs arise because of relationship specific investments and asset specificity. Assets are specific to a transaction when they have great value within the context of a particular transaction but little value outside it. This leads to the possibility of opportunism. Insofar as contracts are incomplete, as uncertainty unfolds and unanticipated events occur, the contract will need to be renegotiated and if one party has made a sunk investment via the development of assets specific to the relationship then the other party could attempt to opportunistically appropriate an undue part of the payoffs to this investment by threatening to withdraw from the relationship. Significantly Gibbon sees the adaptation interpretation as being independent of these hold-up type issues, which makes the adaptation view a non-standard interpretation of Williamson’s work.

Williamson can be seen as proposing that both asset specificity and adaptation are necessary if a theory of the firm is to be realistic.

What Foss, Lando and Thomsen refer to as the ‘firm as an ownership unit’ approach to the firm is the property rights theory or incomplete contracts theory of the firm due to Grossman and Hart (1986, 1987), Hart and Moore (1990) and Hart (1995). The central idea in the property rights approach is that as contracts are incomplete the allocation of control rights affects the incentives that people face and thus their behaviour and the allocation of resources. This theory defines ownership of an asset as the possession of the *residual* control rights over that asset.⁵⁹ A firm is defined as a collection of jointly-owned assets. This means that the distinction between an independent contractor and an employee turns on who owns the non-human assets with which the agent works. An independent contractor owns his own tools while an employee does not.

The importance of ownership is that if a non-contractible, specific to an asset, investment is undertaken then a non-owner risks being held-up by the owner. Thus the property rights theory would say that whoever makes the most important, non-contractible, asset-specific investment should be the owner of the asset. The optimal size of a firm must balance two opposing forces: hiring an employee means hiring someone who lacks optimal incentives since they risk being held up by the firm because they can be fired, thereby separating them from the assets they need to be productive, versus using an independent contractor who could hold-up the firm by threatening to quit the relationship and taking his assets with him.

This trade-off gives a theory which can determine the boundaries of the firm.

In many cases it is difficult, if not impossible, to write complete state-contingent contracts. In such circumstances people will often rely on what has been referred to as ‘unwritten codes of conduct’, that is, implicit contracts. The underlying idea in the implicit contract theory of the firm is that there are differences in the way implicit contracts function between firms (‘rational contracting’) and within firms (an employment agreement).

Baker, Gibbons and Murphy (2002) (BGM) make this point that implicit contracts occur both within and between firms and argue that the difference between them lies in what happens if the implicit contract breaks down. An independent contractor can leave the relationship and take the assets belonging to it with him. This an employee can not do. In BGM an independent contractor can, if he wants, sell the finished product elsewhere while an employee does not own the finished product and thus can not leave the relationship with the asset or the product. The strength of the threat to discontinue the relationship determines the implementability of implicit contracts. As an example consider the situation where the market for the good is highly volatile. In this case a relational contract may be unworkable since the supplier has an incentive to violate the implicit contract when the market price is high. If the supplier is part of the firm such an option does not exist and the implicit contract that holds the internal transfer ‘price’ constant may be

⁵⁸Williamson (1975), for example, makes use of Simon (1951) in Chapters 4 and 5.

⁵⁹Residual control rights are those rights associated with being able to use the asset under conditions not specified in the contract.

self-enforcing.⁶⁰ The implicit contracting theory can be seen as being related to Williamson's idea that the resolution of disputes is more easily achieved within firms than between firms in the sense that mechanisms for dispute resolution can be seen as a feature of a system of self-enforcing implicit contracting within the firm.

The last of the subgroups identified by Foss, Lando and Thomsen is the 'the firms as a communication-hierarchy' subgroup. Here the firm is viewed as a communications network designed to minimise both the cost of processing new information and the costs associated with the dispersing information among the members of the firm. Clearly communication is costly in that it takes time for people to absorb new information that they have been sent. But this time can be reduced by having particular agents specialising in the processing of particular types of information. In the model from Bolton and Dewatripont (1994), for example, each agent handles a particular type of information with the different types being aggregated via the communications network. Teams, firm-like structures, arise when the benefits to specialisation are greater than the costs of communication.

The major problem with this approach to the firm is that it can't explain the boundaries of the firm. The theory does not explain why communication hierarchies can exist within firms but not between firms.

We add one additional subgroup to the Foss, Lando and Thomsen incomplete contracting classification: the reference point theory.⁶¹ This approach arose, in part, from the Maskin and Tirole (1999) critique of theory of incomplete contracts.⁶² Maskin and Tirole argue that information which is observable to the contracting parties but not to a third party, e.g. the courts – such, so-called, non-verifiable information is normally assumed to be the reason for contractual incompleteness in the 'firm as an ownership unit' approach – can be made verifiable to the third party by the use of ingenious revelation mechanisms. The contracting parties write into their contract a game which when played gives the appropriate incentives for them to truthfully reveal their private information in equilibrium. This undermines the non-verifiability approach to incomplete contracts. To deal with this Hart and Moore (2008) developed the reference point approach to contracts. The basics of the application of the reference point theory to theory of the firm can be seen in the model of Hart and Moore (2007) (HM). We deal with this paper in some detail as the reference point literature is still relatively new and thus is not as well known as the theories considered up to this point.⁶³

The basic ideas underlying the reference point theory can be outlined as follows. Consider a situation where a buyer B wants a good from a seller at some future date 1. Assume the good is a homogeneous widget. Also assume that there is a 'fundamental transformation' between dates 0 and 1, that is, what starts as a situation of perfect competition at date 0 evolves into one of bilateral monopoly by date 1. The parties meet and contract at date 0 but there is uncertainty about the state of the world at this time. This uncertainty is resolved shortly before date 1. There is symmetric information thought-out but the state of the world is not verifiable. A date 0 contract can be thought of as specifying a set of possible price-quantity pairs which form the set of possible outcomes of B and S 's date 1 transaction. Note that the outcomes cannot be state contingent since the state itself is not verifiable. A mechanism for choosing from among the set of possible outcomes may also be included as part of the date 0 contract.

Importantly for the HM story, the date 0 contract acts of a reference point for the contracting parties feelings of entitlement at date 1. Neither party feels entitled to an outcome not included in the contract. The contract is seen as "fair" since it was negotiated under the competitive conditions prevailing at date 0.⁶⁴ Problems can arise, however, when choosing among the different outcomes allowed under the contract. HM suppose that each party feels entitled to the best pos-

⁶⁰Self-enforcing here means that each party lives up to the other party's expectation in fear of retaliation and the breaking down of cooperation.

⁶¹Hart (2008) offers an intuitive introduction to the reference point approach to the theory of contracts.

⁶²See Maskin (2002) and Aghion and Holden (2011: 190-3) for non-technical discussions.

⁶³For a more detailed discussion of the reference point literature see Walker (forthcoming).

⁶⁴This is an important point for all the papers in the reference point approach. See Fehr, Hart and Zehnder (2011) for experimental evidence related to this issue.

sible outcome allowed under the contract.⁶⁵ This means that it is likely that at least one of the parties, if not both, will feel disappointed or aggrieved by the actual outcome.

A second important assumption built into the HM theory is that any outcome is not perfectly contractible even at date 1. Each party has the freedom to choose between perfunctory performance and consummate performance but only perfunctory can be contracted on. Consummate performance will be provided if the party feels well treated but each party will ‘shade’ if they feel aggrieved.⁶⁶

To make matters a little more precise, suppose that if the outcome that is chosen from those available under the contract causes S to feel aggrieved by $\$k$, that is, S ’s actual payment is $\$k$ less than the best possible outcome, then S will shade on her performance to such a degree that B ’s payoff falls by θk . θ being an exogenously given parameter where $0 < \theta \leq 1$. A similar situation with regard to shading pertains to B . There a symmetry here, both B and S can shade and θ is the same for both. Shading can not occur if there is no trade.

Assume further that B ’s value of the widget at date 1 is v and that S ’s cost of production is zero but there is an opportunity cost of r . That is, trade with B means S foregoes an alternative income of r . Thus, trade is efficient if and only if $v \geq r$. At date 0, v and r are random variables with a probability distribution which is common knowledge. Also no third party is able to tell who is at fault if trade does not take place at date 1. This means that trade is voluntary. Given these assumptions Hart and Moore (2008) are able to show that it is only the difference between the trade price and the no-trade price that matters and that is possible to normalise the no-trade price to be zero. HM also assume that lump-sum transfers can be used to carryout any redistribution of surplus at date 0.

The simplest case to consider is that where there is no uncertainty as to the value of v and r , that is, they are constants, and $v > r$. In this situation the first-best can be achieved. All that has to happen is that the parties agree, at date 0, that S will supply the widget to B at date 1 for a given price, p , where $r < p < v$. This contract would ensure trade and would result in no aggrievement because both parties receive the best outcome permitted under the contract. Note that the contract only specifies one outcome, trade at price p .

While this form of contract achieves the first-best not all contracts, even in this no uncertainty world, do so. For example, consider a contract that specifies that the trade price can be anything in the range $[r, v]$ and that B will choose the price. Here B will choose the lowest price possible, $p = r$, at date 1. Note however that this will cause S to be aggrieved since the best possible price for her, $p = v$, was not chosen and thus she will shade resulting in a deadweight loss of $\theta(v - r)$.⁶⁷

Things are more interesting, however, if v and r are uncertain. In this case any contract which specifies a single trading price, p , will ensure trade if and only if $v \geq p \geq r$, that is, if and only if both parties gain from trade. The problem is that as v and r are now stochastic, it can not be guaranteed that it is possible to find a single p that lies between v and r whenever $v > r$. HM point out that under such conditions a contract that specifies a range of trading prices $[p, \bar{p}]$ can be superior to a single price contract. Hart and Moore (2008) show that it is not necessary to go beyond a contract which specifies a no-trade price of zero, as above, a trading price range of $[p, \bar{p}]$ and lets B choose the price. The advantage of the large price range is that it makes it more likely that B can find a price between v and r whenever $v \geq r$. The cost is that there are typically many feasible prices between v and r when $v \geq r$ and B will pick the lowest price. This means that S will feel aggrieved that B did not pick the highest price and will therefore shade, resulting in a deadweight loss. The optimal contract will trade off these two effects.

Thus far one important issue has been ignored. If $v > r$ but there is no price in the range $[p, \bar{p}]$ such that $v \geq p \geq r$, it would be expected that the parties would renegotiate their contract. But renegotiation doesn’t change the analysis in any fundamental ways as is shown in Hart and Moore (2008).

⁶⁵Such an extreme assumption not necessary for the analysis but does simplify the workings.

⁶⁶Shading is where a party performs at the lower, perfunctory, level rather than the higher, consummate, level.

⁶⁷ S ’s best outcome is $p = v$ while the actual outcome is $p = r$ and thus S is aggrieved by the amount $v - r$ which means she shades, thereby lowering B ’s payoff, by an amount $\theta(v - r)$. This is the deadweight loss. B does not shade as she receives her best outcome.

Next HM turn to the issue of ownership. Up to this point HM have implicitly assumed that B and S are separate entities, that is, they are “nonintegrated”. Now suppose that B acquires S ’s firm (S ’s non-human assets) at date zero. This is interpreted as “integration”. It amounts to saying that B now owns and possesses the widget. HM take this to mean that B can get someone other than S to produce the widget, at zero cost, at date 1. It is assumed that S ’s human capital, along with the widget, is still needed to realise the opportunity cost, r .⁶⁸ Effectively for S to earn r , B must sell the widget back to S . If no trade occurs B earns v since she already owns the widget, while if trade does occur, S earns r but pays p . Trade is now efficient if and only if $r \geq v$; trade is still voluntary. In this situation a contract consists of a zero no-trade price and a range of trading prices $[p, \bar{p}]$, with S choosing the price. S will choose the smallest price such that $r \geq p \geq v$, whenever $r > v$.

In place of a complete analysis of nonintegration versus integration HM makes a number of observations on the difference between them. Assume that $v > r$ with probability 1. Then as was noted above, it may be impossible to achieve the first-best. The reason being that in order to ensure trade with probability 1 it may be necessary to have a range of trading prices, but this results in aggrievement and shading whenever there is more than one price satisfying $v \geq p \geq r$. On the other hand integration can achieve first-best because the status-quo has been transformed into one where B owns the widget, which is the efficient outcome. S is irrelevant and does not or cannot shade.

The situation is reversed if $v < r$ with probability 1: now integration is inefficient as a range of prices is required to ensure that B trades the widget to S . This results in aggrievement and thus shading whenever there is a number of feasible prices in the range while for nonintegration the status-quo point has S possessing the widget which is efficient and does not give rise to shading.

The HM model can be thought of as capturing the idea that integration is useful for ensuring input supply in an uncertain world. When $v > r$ but v and r vary, nonintegration is usually inefficient, that is, either trade will not take place when it should or there will be shading, while integration results in the first-best outcome.

The papers making up the reference point approach to the firm have demonstrated that the trade-off between contractual rigidity and flexibility has important implications for the organisation of firms. The unifying feature shared by the reference point papers is the application of the idea that contracts act as a reference point for feelings of aggrievement and thus acts of shading. The Hart and Moore (2008) theory and its extensions provide an explanation for the existence of long-term contracts in the absence of relationship-specific investments, which are assumed in most of the incomplete contracts approaches to the firm. Also the reference point theory can shed new light on the roles of the employment relationship and authority. In work extending the theory, Hart (2009) reintroduces assets into the model and shows that previously hard to explain observations in the empirical literature on contracting and integration can be explained by the reference point approach. Hart and Holmstrom (2010) offers a theory of firm scope. They provide an analysis that moves the focus of the theory away from the role of non-human assets in determining a firm’s boundaries towards a theory where the activities undertaken by the firm determine the firm’s scope.

Another recent contribution to the theory of the firm, but one which can be seen as being outside – even if related to – the post-1970 mainstream, and thus a contribution that does not fit neatly into the Foss, Lando and Thomsen classification, is Spulber (2009). If we think of the mainstream theory of the firm as being concerned with three basic topics to do with the existence, boundaries and internal organisation of the firm then Spulber’s book is somewhat outside of the mainstream but the issues it deals with are related closely enough to those of the mainstream to justify a brief overview.⁶⁹ Spulber’s framework offers insights into a number of important issues missing from the mainstream and thus lays the groundwork for future research in both the theory of the firm and industrial organisation more generally. The three mainstream topics are mentioned in Spulber’s book but they are not the main focus of his analysis. For Spulber “*The*

⁶⁸ S gets zero if she has only their own human capital to work with. To get r , S requires both her human and non-human assets (or equivalently the widget).

⁶⁹For criticisms of the Spulber approach to the firm from inside the mainstream see Hart (2011b).

Theory of the Firm seeks to explain (1) why firms exist, (2) how firms are established, and (3) what firms contribute to the economy". (Spulber 2009: ix). Particular issues which are a focus of Spulber's framework but with which the mainstream does not deal, or at least does not deal with fully, include a theory of the entrepreneur and a theory of market creation. Spulber's book is an attempt to create a general approach to microeconomics in which entrepreneurs, firms, markets, and organizations are all endogenous. This makes his intended contribution wider than that of the mainstream theory of the firm.

To start Spulber defines a firm "[...] to be a transaction institution whose objectives differ from those of its owners. This separation is the key difference between the firm and direct exchange between consumers". (Spulber 2009: 63). Note that under this definition organizations such as clubs, basic partnerships, many family firms, worker cooperatives, non-for-profit organisations and public enterprises are not firms. The fundamental reason being that the objectives of these types of organizations cannot be separated from those of their owners. This separation of the objectives of owners and the firm also justifies the profit maximisation objective. Consumers who are the owners of the firm, obtain income via the firm's profits and thus want profit maximisation so they can maximise their consumption (utility).

Spulber has three additional players in his story: consumers, organisations and markets. Consumers are individuals who consume the goods and services generated within the economy. Organisations are transaction institutions whose objectives cannot be separated from those of their owners. Markets are transaction mechanisms that bring buyers and sellers together.

For Spulber the entrepreneur is important because it is their efforts that leads to the creation of firms. The interaction of available market opportunities and the individual's preferences, endowments and other such characteristics leads the consumer to take on the role of entrepreneur. The individual is an entrepreneur for the time it takes to establish the firm. Assuming that the firm is successfully established, the entrepreneur's role changes to that of an owner of the firm. Ownership is valuable insofar as it provides returns to the (former) entrepreneur. This change from entrepreneur to owner is in Spulber's terms, 'the fundamental shift'. (Spulber 2009: 152). The important point is that before the change from entrepreneur to owner, the objectives of the organisation cannot be separated from those of the (then) entrepreneur. The start-up enterprise is not a fully formed firm because of this. After the fundamental shift has occurred, the entrepreneur is now the owner and the firm's objectives are separate from those of the owner, which means a firm has now been fully established.

In addition to firms other organisations can be formed that allow consumers to take advantage of joint production. Such advantages include economies of scale, public goods, common property resources and externalities. Unlike the firm, the objectives of the consumer organisation reflect the consumption objectives of the members of the organisation. A problem with a consumer organisation is that it can experience inefficiencies due to free riding. Firms overcome such problems by separating the objectives of the organisation from the consumption objectives of its owners and thereby inducing profit maximisation. Profit maximisation may not, however, achieve full efficiency due to problems such as allocative inefficiencies that result from market power. Thus the comparison between a firm and a consumer organisation depends on the trade-off between profit maximisation and free riding. When the number of consumers is small, free riding problems tend to be small and thus a consumer organisation may be more efficient than the firm.

Given we have firms, firms can create markets. Firms can act as intermediaries and in doing so they increase the gains from trade and reduce transactions costs when consumers are separated by time, distance and uncertainty. Firms create markets by providing centralised mechanisms for matching consumers – in their roles as buyers and sellers – in a more efficient manner than decentralised exchange can achieve. Market making firms are required to buy and sell at any moment meaning that buyers and sellers avoid the costs involved with trading delays and the risk of being unable to find a trading partner. In the world of financial markets, market-making firms provide liquidity by being ready to buy and sell financial assets. Addition transaction efficiencies are bought about by market-making firms being able to consolidate trades which allows traders to reduce the costs involved with having to find multiple trading partners or with having

mismatched trades. Firm can also avoid the problems inherent with complex bartering arrangements by simplifying the trading process via the use of posted prices. Supply and demand can be equated in a market by firms adjust their buying and selling behaviour thereby reducing potential losses due to market imbalances. Thus market creation is endogenous.

Spulber offers insights into a number of issues that the mainstream theory of the firm does not deal with well, if at all. The discussion of issues such as the role of the entrepreneur and the creation of markets are important issues that lie outside of the mainstream of the theory of the firms, at least as far as the mainstream is conceived of here. Such issues do however raise questions for the future of the theory of the firm and the theory of industrial organisation.

The post-1970 theories of privatisation

An often neglected topic in the theory of the firm is that of state owned firms and their privatisation.⁷⁰ But as Hart (2003: C69) makes clear there are close parallels between the theories of the firm and of privatisation.

“Let me begin by discussing the very close parallel between the theory of the firm and the theory of privatisation. In the vertical integration literature one considers two firms, A and B. A might be a car manufacturer and B might supply car-body parts. Suppose that there is some reason for A and B to have a long-term relationship (e.g., A or B must make a relationship-specific investment). Then there are two principal ways in which this relationship can be conducted. A and B can have an arms-length contract, but remain as independent firms; or A and B can merge and carry out the transaction within a single firm. The analogous question in the privatisation literature is the following. Suppose A represents the government and B represents a firm supplying the government or society with some service. B could be an electricity company (supplying consumers) or a prison (incarcerating criminals). Then again, there are two principal ways in which this relationship can be conducted. A and B can have a contract, with B remaining as a private firm, or the government can buy (nationalise) B”.

and

“[. . .] the issues of vertical integration and privatisation have much more in common than not. Both are concerned with whether it is better to regulate a relationship via an arms-length contract or via a transfer of ownership”. Hart (2003: C70)

The incomplete contracting framework discussed in the previous subsection gives an approach which can be utilised to study the difference between public and private ownership. In fact incomplete contracts are a necessary condition to explain the differences between the two forms of ownership. In a world of complete or comprehensive contracts there is no difference between private and state owned firms. In both cases the government can write a contract with the firm that will anticipate all future contingencies - it will detail the managers' compensation, the pricing policy of the firm, how changes in technology will change the firm's products etc - and thus the outcome under both forms of ownership will be the same.

This intuition has been formalised into a series of Neutrality Theorems. These theorems establish the conditions under which private or public ownership of productive assets is irrelevant for the final allocation of resources. Consider first the 'fundamental privatisation theorem' due to Sappington and Stiglitz (1987). Assume the government's aim is to simultaneously achieve three objectives: (i) economic efficiency; (ii) equity; (iii) rent extraction. What Sappington and Stiglitz show is that the government can design an auction scheme that will result in these three

⁷⁰Hart (1995: 11-2), for example, has a short section entitled 'An omitted topic: public ownership'. Hart writes “[a] very important topic not considered concerns the optimal balance between public and private ownership. [. . .] This issue has always been a central one in the economic and political debate, but it has attracted new attention in the last few years as major industries have been privatized in the West and the socialist regimes in Eastern Europe and the former Soviet Union have dissolved”.

objectives being achieved and where both public and private production give the same outcome. The government has a 'social' valuation of the level of output. This valuation embodies the government's concerns with regard to equity issue such as the consumption levels of the good among different classes of citizens. It is assumed that the costs of production are such that production by a single firm is optimal but there are at least two risk-neutral firms, who have symmetric beliefs about the least-cost production technology, willing to bid to be the supplier. The government auctions off right to the supplier of the good with the understanding that the supplier receives a payment which equals the social evaluation. The most efficient firm will win the contract with the highest bid, which will equal the firm's (expected) profits, and will set the production level most preferred by the government. Rent extraction is achieved since the winning bid equals the firm's profits and economic efficiency is achieved since the most efficient firm is selected as the producer and the firm produces the government's preferred (social welfare maximising) level of output.

A simple example of this mechanism is given by Bös (1991: 20). Let the payment received by the firm equal the government's social valuation which equals the sum of consumer surplus plus revenue.⁷¹ This induces a profit maximising firm to maximise the sum of consumer and producer surplus. This implies technological and allocative efficiency. Since the highest offer in the competitive auction is identical to the expected profit of the firm, the expected monopoly profit goes to the government.

Shapiro and Willig (1990) obtain a similar result for a setting in which a public-spirited social planner or framer decides on the nationalisation/privatisation outcome and sets up the governance structure for the enterprise chosen. The framer's decision is driven by the informational differences between private and public ownership. The important pieces of information are: (i) information about external social benefits generated by the firm; (ii) information concerning the difference between the "public interest" and the private agenda of the regulator; (iii) information about the firm's profit level (cost and demand information).

First consider the case where the firm is state owned. Here the firm is run by a public official that Shapiro and Willig refer to as a Minister. By virtue of his role in managing the enterprise, the minister receives the private information about the profitability of the enterprise. By virtue of his position in the public sector, the minister also observes information that bears on the external social benefits generated by the enterprise's operations. Given this information the minister makes decision as to the level of investment in the firm and the level of output for firm. The overall social welfare function that the framer seeks to maximise is the sum of external benefits plus enterprise profits where there is a magnification factor added to the profit term which equals the unit cost of raising public funds, including any distortions caused by the taxes required to finance public sector operations. The minister's objective function is that of the framer plus a term related to the private agenda of the minister where there is a weighting parameter attached to private agenda term which measures how easily the minister can extract these benefits. This parameter can be interpreted as being a proxy for how well the political system works. The better the system the greater the limits on what the minister can extract.

If the firm is a private company then it is managed by a professional manager and is overseen by a regulator. The manager observes the profitability of the firm while the regulator learns the nature of the externality variable and the private agenda variable. The regulator designs a regulatory scheme that offers the expectation of a competitive rate of return on the private firm's sunk capital. The firm then maximises profit subject to the regulatory scheme while the regulator has the same objective function as the minister under state ownership. The framer's objective is to maximise the sum of the external benefits plus profits net of the cost of raising any public funds needed to make the transfers to the private company required under the regulatory scheme.

The important difference between the two ownership forms is who receives the information about cost and demand conditions. The manager is the informed party under private ownership while under public ownership the minister is informed. This means that an informational barrier is created between the firm and the government by privatisation. The advantage of this barrier is that it reduces the discretion the minister has to interfere with the working of the firm. The

⁷¹This is the total area under the demand curve for a given quantity.

disadvantage is that it makes it more difficult for the regulator to motivate the firm to pursue social welfare objectives.

When considering neutrality results first consider the operation of an enterprise in an environment in which there is no private information whatsoever. Suppose all information about the external benefits of the enterprise and all information about its profitability is contractible. In such circumstances, the regulator could put in place a set of taxes or subsidies, contingent on what will become commonly known realisations of the public costs and benefits of the enterprise's operations. These taxes and subsidies could be designed to induce the owners to operate the enterprise to serve precisely the regulator's objectives in every contingency.

Perhaps it is not surprising that one can obtain a neutrality result in the complete absence of noncontractible private information, for in such a case there is no truly active role for the managers of the enterprise. They need only carry out the detailed instructions left by the minister or the regulator, and the manager cannot claim that there will ever be any new information or extenuating circumstances that can justify departures from that mechanical mandate.

The more interesting neutrality results arise in situations where there is private information. First assume that the private information about the firm's profitability is known only after the investment is made but the private information concerning public impacts and private agenda is known to the regulator when he must commit himself to the regulatory mechanism, before the time of the investment decision. Under these conditions, the regulator can exert sufficient indirect control over the private firm to obtain the same outcome and payoff as under public ownership, so the framer is indifferent between public and private enterprise. The regulator's control is secured by paying the firm according to the schedule which takes into account the sum of external benefits generated plus the private agenda of the regulator plus the smallest possible payment that will induce the firm to invest. With this schedule, the regulator induces the same actions and achieves the same payoffs as does the minister under public enterprise. The mechanism operates by forcing the firm to internalise the objectives of the regulator.

The second distinct case occurs when private information concerning both costs and public impacts is revealed only after the investment commitment must be made. Only the prior probability distributions of the private information of the regulator and the profitability of the firm are known at the time the investment decision must be effected. After the investment has been made, but before the activity level must be chosen, the private information of the regulator will become known to him and the nature of the firm's profitability will be revealed to the manager of the enterprise. Again, the regulator's optimal payment scheme results in the same choices of activity levels and the same expected drain on the treasury that would be the result of public enterprise. The logic behind this result is a straightforward extension of the analysis of the first case. Here the regulator commits himself to the menu of payment schedules, with the understanding that he will choose a particular schedule from this menu after investment is made and his private information is revealed to him, but still before the activity level must be chosen by the firm. The firm is indifferent, *ex ante*, about which particular schedule will be chosen from the menu by the regulator, because each of them offers the same zero level of expected profits, that is just enough to induce the firm to make the investment. Once the regulator learns his private information, he will be motivated to select the payment schedule corresponding to that information because that schedule is optimal for his objective function. Given this payment schedule, the firm will be motivated to choose the same activity level as in the first case above, and here too that is the optimum from the perspective of either the regulator or the public minister.

In the third case of neutrality the private firm has private information about its costs before the investment decision must be made. It is assumed that there are no costs to raising public funds and thus any transfers from the treasury are not a matter for concern to the framer, the regulator or the public minister. Because the firm knows information about its profitability and the regulator is aware of that fact but does not know this information himself, the regulator, to assure that investment will be made, must commit to a payment schedule or to a menu of schedules that provides non-negative profit for all demand/cost cases. Here, because of the stipulation that public funds can be raised at zero cost, this requirement poses no problem for the regulator: he is perfectly willing to add enough funds to any payment schedule to assure

its profitability in the light of his indifference to transfers from the treasury. Consequently, it is optimal for the regulator to offer the firm internalisation schedules, each with different levels of investment funds, such that these funds are sufficiently large to guarantee the firm non-negative profit even if its profitability level is the worst possible. In the end, the regulated firm chooses the same activity levels that the public enterprise would choose, but the drain on the treasury caused by regulation is greater than that caused by public enterprise. Since, in this case, however, that drain is not a matter of concern, the framer would find no difference between the performance of public and private forms of organisation.

The third neutrality result is that of Shleifer and Vishny (1994). Their starting point is the idea that politicians control SOEs in order to achieve political objectives, such as excess employment and/or high wages. In this model the politician derives benefits from this inefficient allocation of resources, as they create political support for him. If the firm is privatised then the politician must bargain with the manager of the firm to get the outcome he wants. Clearly the manager, who aims to maximise profits, and the politician, who wants political support, have conflicting objectives. The firm will not want to expand employment above the profit maximising level as the politician wishes to do. The politician must make a transfer, from the treasury, to the firm to induce the taking on of the extra workers. This is a problem to the politician since the transfer is costly to him as taxes need to be raised to finance the subsidy.

The Shleifer and Vishny model allows for a complete separation of income rights and control rights. There is no clear-cut dichotomy between state-owned and private firms in the model as it allows for four corporate forms:

- i a SOE, the Treasury has income rights and the politician has control rights;
- ii a regulated firm, the private owners have income rights, but the politician has control rights and can interfere in the operating activity of the firm;
- iii a 'corporatised' firm, when the government has income rights, but the control rights are in the hand of the firm's management;
- iiii a purely private firm, when the manager/owner has both income and control rights.

As the model has the two parties bargaining, disagreement points have to be identified. These points are where the politician and the manager control the firm. When the politician controls the firm he has control over the manager and is able to bring the firm down to zero profits. He can use the firm's cash flow to hire extra labour up to the point where the marginal benefits of the excess employment equals the marginal cost of raising public funds. Under control by the manager, the manager has power over the politician, and the firm produces at the efficient level (with zero excess labour) but does not receive any transfer from the Treasury.

As far as the manager and the politician are concerned, the efficient point is reached when the level of excess employment reaches the point where the marginal political benefits equals the wage, which is the marginal cost of labour. At this point the amount of excess labour employed is lower than that under politician control and the subsidy paid to the firm is higher than under private control.

The neutrality result that Shleifer and Vishny present is basically an application of the Coase Theorem to privatisation. As side payments are allowed - or more correctly in this case, when the manager and politician can freely bribe each other - then the manager and the politician will reach the jointly efficient solution no matter what the initial allocation of income and control rights.

The importance of the above theorems is that they outline the conditions under which ownership of the firm does not matter. Of all the assumptions on which the irrelevance results hinge the most important requirement is that complete contingent long-term contracts can be written and enforced. But writing complete contracts is only possible in a world of zero transaction costs. In a positive transaction costs world only incomplete contracts can be written but contractual incompleteness creates a role for ownership - making decisions under conditions not covered in the contract. It is only within such an environment that we can explain why privatisation matters,

that is, why the behaviour of state owned and private companies differ. This reliance on incomplete contracts means that the theory of privatisation can be seen as forming a seventh subgroup of the property rights group outline above.

These results also shows why the previous theoretical privatisation literature was largely unsuccessful.⁷² That literature took a 'complete' or 'comprehensive' contracting perspective, in which any imperfections present in contracts arose solely because of moral hazard or asymmetric information. But as Hart (2003: C70) notes

“[. . .] if the only imperfections in are those arising from moral hazard or asymmetric information, organisational form - including ownership and firm boundaries - does not matter: an owner has no special power or rights since everything is specified in an initial contract (at least among the things that can ever be specified). In contrast, ownership does matter when contracts are incomplete: the owner of an asset or firm can then make all decisions concerning the asset or firm that are not included in an initial contract (the owner has 'residual control rights').

Applying this insight to the privatisation context yields the conclusion that in a complete contracting world the government does not need to own a firm to control its behaviour: any goals - economic or otherwise - can be achieved via a detailed initial contract. However, if contracts are incomplete, as they are in practice, there is a case for the government to own an electricity company or prison since ownership gives the government special powers in the form of residual control rights”.

Thus privatisation matters only in an incomplete contracts world. In such an environment the allocation of residual control rights will differ and so the behaviour of publicly owned firms will differ from that of privately owned firms and thus ownership and therefore privatisation will become meaningful.

Schmidt (1996a)⁷³ considers a monopolistic firm that produces a public good in a world of incomplete contracts. His model is multiple period with the privatisation decision being made in the initial period. That is, the government must decide whether to sell the SOE to a private owner-manager or keep it in state hands and hire a professional manager to run it. Importantly knowledge concerning the firm's cost is private information known only by the firm's owner. Given this, privatisation amounts to a transfer of private information from the government to the private owner. In the next period the manager selects his effort level and the state of the world is then revealed. The importance of the manager's effort level is that it affects the probability of the state of the world. A high level of effort from the manager results in productive efficiency being enhanced and costs being lowered for any level of output. In the last period, the government selects the transfer scheme and payoffs are revealed.

When the firm is an SOE the government observes the firm's realised cost function and thus can implement the first-best allocation by choosing the ex post efficient level of production. But the manager's wage will be fixed, since contingent contracts can not be written, and thus independent of level of output. Given this the manager has no incentive to exert effort and the government knowing this will therefore offer him only his reservation wage.

On the other hand when the firm is in private hands the government does not know the exact cost structure of the firm. In an effort to get the private owner to produce the efficient level of output the government must provide an incentive via the payment of an informational rent. But if transfer are costly it will be impossible to implement the optimal allocation and therefore the cost to private ownership is an inefficiently low level of production. However given the rent payment provides an incentive to increase effort, productive efficiency is greater.

Schmidt's main conclusion is therefore that when the monopolistic firm produces a good or service which provides a social benefit, there is a trade-off between allocative and productive efficiency that needs to be considered when deciding if a firm is to be privatised. The equilibrium

⁷²See Bös (1991) and Vickers and Yarrow (1988) for discussions of this literature.

⁷³Schmidt (1996a) is variant of Schmidt (1996b). 1996b considers the case of privatisation to an employee manager while 1996a applies to the case of privatisation to an owner-manager. While this second case is less realistic it is simpler and does not require the assumption that the manager is an empire builder that is utilised in 1996b.

production level is socially suboptimal but the incentive for better management results in cost savings. Considered overall the welfare effect of privatisation should be positive for cases where the social benefits are small, but social welfare will be greater under public ownership for those cases where production exhibits large social benefits.

An important implication of this is that a case can be made for privatisation even when the government is a fully benevolent dictator who wishes to maximise social welfare. Even if all the deficiencies of the political system could be remedied it is still possible for privatisation to be superior to state ownership.

In the Laffont and Tirole (1991) model a firm is assumed to be producing a public good with a technology that requires investment by the firm's manager. In the case of a public firm this investment can be diverted by the government to serve social ends. For example, the return on investment in a network could be reduced by the government if it were to allow ex post access to the general population. Such an action may be socially optimal but would expropriate part of the firm's investment. A rational expectation of such an expropriation would reduce the incentives of a public firm's manager to make the required investment. For a private firm, the manager's incentives to invest are better given that both the firm's owners and the manager are interested in profit maximisation. The cost of private ownership is that the firm must deal with two masters who have conflicting objectives: shareholders wish to maximise profits while the government pursues economic efficiency. Both groups have incomplete knowledge about the firm's cost structure and have to offer incentive schemes to induce the manager to act in accordance with their interests. Obviously the game here is a multi-principal game which dilutes the incentives and yields low-powered managerial incentive schemes and low managerial rents.⁷⁴ Each principal fails to internalise the effects of contracting on the other principal and provides socially too few incentives to the firm's management. The added incentive for the managers of a private firm to invest is countered by the low powered managerial incentive schemes that the private firm's managers face. The net effect of these two insights is ambiguous with regard to the relative cost efficiency of the public and private firms. Laffont and Tirole can not identify conditions under which privatisation is better than state ownership.

In the Shapiro and Willig paper discussed above privatisation is considered in a context where the regulator pursues a different agenda from the framer. Assume that either information about profitability is known before investment is decided upon or that there are costs to raising public funds. In these cases the neutrality results of Shapiro and Willig don't hold. The equilibrium behaviour of the minister who is in charge of the firm is virtually unconstrained and he will set the activity levels of the firm as to maximise his utility. The regulator of the private firm has a more complex problem to deal with. This involves the designing of regulatory scheme which ensures non-negative profits for the firm. Given this is a case of optimal regulation under asymmetric information we would expect to see the firm enjoying informational rent, which are proportional to the activity chosen. As public funds are costly to raise these transfers are costly to the state.

The trade-off in this model is driven by how easily the public official can interfere with the operations of the firm. If the public official's objectives are the same of the (welfare maximising) framer, i.e. the public official has not private agenda, then public ownership is optimal. In this case private ownership reduces performance since the firm extracts a positive information rent. But when there is a private agenda then a reduction in discretion may increase welfare. Politicians find it easier to distort the operations of a firm in their favour when that firm is an SOE and under the direct control of the minister. The regulated private firms does earn a positive rent but is less subject to the control of the regulator. This means that regulated private firms are likely to out perform SOEs in poorly functioning political systems, which are open to abuse by the minister, and where the private information about the profitability of the firm is less significant. This makes it easier for the regulator to get the firm to maximise social welfare.

In Boycko, Shleifer and Vishny (1996) information problems do not explain the difference between public and private firms. Here it is differences in the costs to a politician of interfering

⁷⁴Technically the multi-principal distortion is similar to the double marginalisation on two complementary goods sold by noncooperative monopolists.

in the activities of the different types of firms that explains the effects of privatisation. The starting point of the paper is the observation that public firms are inefficient because they address objectives of politicians rather than maximise efficiency. One common objective for a politician is employment. Maintaining employment helps the politician maintain his power base. In their model Boycko, Shleifer and Vishny assume a spending politician, who controls a public firm, forces it to spend too much on employment. The politician does not fully internalise the cost of the profits foregone by the Treasury and by the private shareholders that the firm might have.

Boycko, Shleifer and Vishny argue privatisation can be a strategy to reduce this inefficiency in state-owned enterprises. By privatisation they mean the reallocation of control rights over employment from politicians to a firm's managers and the reallocation of income rights to the firm's managers and private owners. The spending politician will still want to maintain employment and can use government subsidies to 'buy' excess employment at the private firm. In this model the advantage of privatisation is that it increases the political costs to maintaining excess employment. It is less costly for the politician to spend the profits of the state-owned firm on labour without remitting them to the Treasury than it is to generate new subsidies for a privatised firm. Given that voters will be unaware of the potential profits that a state firm is wasting on hiring excess labour they are less likely to object than they are to the use of taxes, which they know they are paying, to subsidise a private firm not to restructure. This difference between the political costs of foregone profits of state firms and of subsidies to private firms is the channel through which privatisation works in this paper.

Shleifer and Vishny (1994) is a continuation of research stated in Boycko, Shleifer and Vishny (1996). As with the 1996 paper Shleifer and Vishny assume that there is a relationship between politicians and firm managers that is governed by incomplete contracts and thus ownership becomes critical in determining resource allocation. As noted above the Shleifer and Vishny model is a game between the public, the politicians and the firm managers. The model derives the implications of bargaining between politicians and managers over what the firms will do. A particular focus is on the role of transfers between the private and state sectors including subsidies to firms and bribes to politicians.

To consider the determinants of privatisation and nationalisation Shleifer and Vishny utilise what they term a "decency constraint" which says that the government cannot openly subsidise a profitable firm. To do so would be seen as politicians enriching their friends. The first, obvious, point made is that politicians are always better off when they have control rights. Control brings political benefits, via excess employment, and bribes, to allow a reduction in the excess employment. Both the Treasury and the politicians prefer nationalisation⁷⁵ to subsidising a money-losing private firm. Control brings bribes and even without bribes politicians get a higher level of employment and lower subsidies when they have control. The Treasury likes the smaller subsidies that come with nationalisation. When it comes to profitable firms politicians like control or Treasury ownership because these firms have a strong incentive to restructure since the profits go to the private owners and they lose little in terms of subsidies due to the decency constraint. To ensure the firms achieve political objectives politicians need control. Given the decency constraint politicians don't want managers who have control rights to also have large income rights since the decency constraint means smaller subsidies are lost if employment is cut and income rights mean the managers gain from restructuring and maximising profits. Politicians who have control prefer higher private and lower Treasury ownership since higher private ownership implies higher bribes. Without bribes the private surplus is extracted via higher levels of employment.

Given that politicians like control, Why would they ever privatise a firm? To explain privatisation the interests of taxpayers must become more prominent. Given this the decision to privatise then becomes the outcome of competition between politicians who benefit from government spending (and bribes) and politicians who benefit from low taxes and support from taxpayers. We would expect privatisation to take place when political benefits of public control are low, and the desire of the Treasury to limit subsidies is high. This is most likely to occur when the political costs of raising taxes to pay subsidies is high and when the political benefits from excess employment are low.

⁷⁵As an SOE the Treasury has income rights and the politician has control rights.

The final paper to be considered is Hart, Shleifer and Vishny (1997). Again in this paper information problems are not the driving force of the analysis of contracting out. The provider of a service, either public or private, can invest his time in improving the quality of the service or reducing the cost of the service. The important assumption is that investments in cost reduction have negative effects on quality. Investments are non-contractible *ex ante*. For the case where the provider is a government employee he must obtain approval from the government to implement any innovation he has created. Given that the government has residual rights the employee will gain only a fraction of return on his investment. This gives him weak incentives to innovate. If the service provider is an independent contractor, i.e. the service has been contracted out, then he will have stronger incentives to both cut costs and improve quality. This is because he keeps the returns to his investment. The downside to private provision is that the incentives to cut costs are strong and the provider does not fully internalise the negative effects on quality of the reductions in cost. With public provision the incentive for excessive cost cutting are reduced as are the incentive for innovation and quality improvements. Costs are always lower under private ownership but quality may be higher or lower under a private owner. Hart, Shleifer and Vishny argue that the case for public provision is generally stronger when (i) non-contractible cost reductions have large deleterious effects on quality; (ii) quality innovations are unimportant; (iii) corruption in government procurement is a severe problem. On the other hand their argument suggests that the case for privatisation is stronger when (i) quality-reducing cost reductions can be controlled through contract or competition; (ii) quality innovations are important; (iii) patronage and powerful unions are a severe problem inside the government.

Reference points, property rights and transaction costs⁷⁶

Among the ‘present’ mainstream approaches to the theory of the firm, the transaction costs and property rights theories are the dominant frameworks. In this section we consider the relationships between these theories and the newer reference point approach. To begin, it has been argued that modelling shading costs is an attempt to model transaction costs. For example, in Hart (2008: 406) it is argued that the shading costs utilised in the reference point approach are akin to “haggling costs” – a kind of transaction cost – and in Hart and Moore (2008: 4-5) it is stated that shading costs can be seen as a shorthand for other kinds of transaction costs, such as rent-seeking, influence, and haggling costs. But just how fully shading costs capture transaction costs and thus how well the reference point approach formalises the transaction cost theory is open to debate.

Williamson (2000: 605-6) argues that one of the most important differences between the property rights approach to the theory of the firm and the transaction cost theory is that the property rights theory introduces inefficiencies at the *ex ante* investment stage while the transaction-cost approach emphasises inefficiencies due to *ex post* haggling and maladaptation.⁷⁷ In the property rights approach there are no *ex post* inefficiencies due to the assumption of common knowledge and costless *ex post* bargaining. The difference is summarised by Gibbons (2010: 283) as:

“[t]he model in question is Grossman and Hart’s (1986) [the property rights model], which explores an alternative to Williamson’s (2000, p. 605) emphasis that “maladaptation in the contract execution interval is the principal source of inefficiency”. Instead, in the Grossman-Hart model, there is zero maladaptation in the contract execution interval, and the sole inefficiency is in endogenous specific investments.

It is striking how different the logic of inefficient investment can be from the logic of inefficient haggling. In their pure forms envisioned here, the two can be seen as complements. For example, the lock-in necessary for Williamson’s focus on inefficient

⁷⁶A previous version of this section appeared in Walker (forthcoming).

⁷⁷Holmström and Roberts (1998: 75-9) and Tadelis and Williamson (2012: 43-6) also discuss the differences between transaction cost economics and the property rights theory. Whinston (2003) looks at the empirical differences between the two theories.

haggling could result from contractible specific investments chosen at efficient levels. But by assuming efficient bargaining and hence zero maladaptation in the contract execution interval, Grossman and Hart focused attention on non-contractible specific investments and hence discovered an important new determinant of the make-or-buy decision: in the Grossman-Hart model, an important benefit of non-integration is that both parties have incentives to invest; in Williamson's argument, an important cost of non-integration is inefficient haggling. In short, the two theories are simply different".

This emphasis on ex post haggling and maladaptation can be interpreted as reflecting a view that internal organisation is better at reconciling the conflicting interest of the parties to a transaction and facilitating adaptation to changing supply and demand conditions when such cost are high.

In so much as contracting in the reference point approach is imperfect even ex post, the reference point theory can be seen as a movement away from the ex ante inefficiencies of the property rights theory and back towards the ex post efficiencies of the transaction cost theory. The imperfect nature of ex post contracting in the reference point approach is as Hart (2008: 294) points out "[...] a significant departure from the standard contracting literature. The literature usually assumes that trade is perfectly enforceable ex post (e.g. by a court of law). Here we are assuming that only perfunctory performance can be enforced: consummate performance is always discretionary", and thus inefficiencies can arise ex post. The development of a tractable model of contracts and organisational form that exhibits ex post inefficiency is one of motivations for advancing the reference point approach in the first place. (Hart and Moore 2008: 4). Hart's interpretation of the reference point theory is that this theory can, in a sense, "[...] be viewed as a "merger" of the transaction cost and property rights literatures". (Hart 2011b: 106).

The fact that the reference point approach does not assume that relationship specific investments are made by the contracting parties is another difference between the property rights theory and the reference point approach. Invoking relationship specific investments is standard in the property rights theory. This is not to say that relationship specific investments can not be introduced into the reference point theory, they can, Hart (2011a) is an example where this is done, but, in general, the reference point theory does not rely on such investment.

Williamson's concept of the "fundamental transformation" is of prime importance for the reference point approach.⁷⁸ The change from an ex ante competitive market to an ex post bilateral setting is what Williamson (1985: 61-3) terms the "fundamental transformation". Hart and Moore argue that such a transformation provides a rationale for the idea that contracts are reference points. "A competitive ex ante market adds objectivity to the terms of the contract because the market defines what each party brings to the relationship. HM assume that the parties perceive a competitive outcome as justified and accept it as a salient reference point". (Fehr, Hart and Zehnder 2009: 562). This is an idea which finds experimental support: see Fehr, Hart and Zehnder (2009), Fehr, Hart and Zehnder (2011) and Hoppe and Schmitz (2011).

But we must also be aware that important features of the transaction-cost theory may still have been left out. How fully shading costs capture the costs of ex post maladaptation and haggling is an open question. When discussing some opportunities for the future of transaction-cost economics, Robert Gibbons (2010: 283) notes that "[...] it may be that Hart and Moore's (2008) "reference points" approach is a productive path. Time will tell [...]". Hart (2011b: 106) concludes "[w]hether this merger [resulting in the reference point theory] will be successful remains to be seen".

⁷⁸Given the importance of fundamental transformation to the analysis of economic organisation Williamson (1985: 63) asks why this notion was ignored for so long. In footnote 23 he gives three reasons: "One explanation is that such transformations do not occur in the context of comprehensive, once-for-all contract—which is a convenient and sometimes productive contracting fiction but imposes inordinate demands on limited rationality. A second reason is that the transformation will not arise in the absence of opportunism—which is a condition that economists have been loath to concede. Third, even if bounded rationality and opportunism are conceded, the fundamental transformation appears only in conjunction with an asset specificity condition, which is a contracting feature that has only recently been explicated".

Summary

A useful classification of the post-1970 mainstream literature is based on seeing the theories forming this literature as being derived from the breaking of either of two assumptions in the standard neoclassical general equilibrium model. One group of theories corresponds to the breaking of the assumption that there are no asymmetries in the information available to contracting parties and thus no principal-agent type problems. The second group of theories violates the assumption that agents can costlessly write contracts. Both these groupings take a Coaseian approach to explaining aspects of the firm insofar as they both set out to answer questions about the existence, boundaries and/or internal organisation of the firm. These were the questions asked, for the first time, in Coase (1937). Famously, part of Coase's answer to these questions was to point out that firms can only exist in a world of positive transaction costs.⁷⁹ This assumption of positive transaction costs separates the current incomplete contracting theories of the firm, including the theory of privatisation, from the neoclassical model of the "firm" which was developed, implicitly, within a zero transaction cost framework. The principal-agent theories can be differentiated from the neoclassical theory by their emphasis on monitoring and incentives issues, which do not arise in the symmetric information neoclassical framework.

Conclusion

While in practice firms may be as old as farming – and even if not that old, "firms" of some description are at least two to three thousand years old – attempts at the formulation of a theoretical explanation for the existence, boundaries and organisation of firms only goes back, at the most, to the 1920s or 1930s, while the current mainstream approach to the theory of the firm is even more recent having been developed only since the 1970s. During the period, roughly, from 1930 to 1970 the mainstream theory of the firm was the neoclassical model in which the firm is seen as a production function or production possibilities set, simply a means of transforming inputs into outputs. Given the available technology, a vector of input prices, and a demand schedule, the firm maximises money profits subject to the constraint that its production plans must be technologically feasible. For the pre-1930 neoclassical period there was no generally accepted theory of the firm. Before this the classical economists had only a theory of aggregate production.

When discussing the neoclassical model of the firm Jensen and Meckling write, "[w]hile the literature of economics is replete with references to the "theory of the firm," the material generally assumed under that heading is not actually a theory of the firm but rather a theory of markets in which firms are important actors". (Jensen and Meckling 1976: 306). The move from a 'theory of markets with firms' to a 'theory of the firm' is the major change that has taken place within the mainstream approach to the theory of the firm over the theory's history.⁸⁰ What we have seen since the 1970s is a movement away from the theory of the firm being seen as developing a component of price theory, namely the component which asks, How does the firm act in factor and product markets?, to the theory being concerned with the firm as a subject in its own right.

Mark Roe (1994: vii) sums up much of this change when he writes,

"[e]conomic theory once treated the firm as a collection of machinery, technology, inventory, workers, and capital. Dump these inputs into a black box, stir them up, and one got outputs of products and profits. Today, theory sees the firm as more, as a management structure. The firm succeeds if managers can successfully coordinate the

⁷⁹The importance of positive transactions costs is the theme that links Coase's two most famous papers. As Demsetz (1996: 565) notes, "The Problem of Social Cost' (Coase, 1960) is R. H. Coase's most cited and most influential work. It is noted for, among other things, demonstrating the importance of incorporating transaction cost into the analysis of externalities and into the analysis of markets more generally. This theme, that markets are not free, is also found in the classic 'The Nature of the Firm' (Coase, 1937), so that, taking the perspective offered by both works, transaction cost turns out to be important whether one is analyzing allocation through the price system or through the firm".

⁸⁰It could be argued that the mainstream theory of the firm has changed while the orthodox (neoclassical) theory has not.

firm's activities; it fails if managers cannot effectively coordinate and match people and inputs to current technologies and markets. At the very top of the firm are the relationships among the firm's shareholders, its directors, and its senior managers. If those relationships are dysfunctional, the firm is more likely to stumble".

In summary, a rudimentary history of the development of the theory of the firm from the 'past' to the 'present' would read: the classical economists had a theory of aggregate production but no theory of firm level production, the neoclassical economists had a theory of firm level production but no theory of the firm, the current literature has started on the task of developing, for the first time, a genuine theory of the firm.

A final point about the models of the firm discussed in this essay is that they highlight a general issue to do with post-1970 microeconomics, that is, the retreat from the use of general equilibrium (GE) models.⁸¹ All the models considered above are partial equilibrium models, but in this regard the theory of the firm is no different from most of the microeconomic theory developed since the 1970s.⁸² Microeconomics such as incentive theory, incomplete contract theory, game theory, industrial organisation etc, has largely turned its back, presumably temporarily, on GE theory and has worked almost exclusively within a partial equilibrium framework.⁸³ This illustrates the point made at the beginning of the paper that there is a close relationship between the economic mainstream and the theory of the firm; when the mainstream forgoes general equilibrium, so does the theory of the firm.⁸⁴

One major path of influence from the mainstream of modern economics to the development of the theory of the firm has been via contract theory. But contract theory is an example of the mainstream's increasing reliance on partial equilibrium modelling. Contract theory grew out of the failures of GE. As Salanié (2005: 2) has argued,

"[t]he theory of contracts has evolved from the failures of general equilibrium theory. In the 1970s several economists settled on a new way to study economic relationships. The idea was to turn away temporarily from general equilibrium models, whose description of the economy is consistent but not realistic enough, and to focus

⁸¹When discussing the influence of Gerard Debreu on economics Duppe (2010: 2-3) nicely sums up the fate of GE as well. "From the point of view of today Debreu's influence on the body of economics could be called zero, in that general equilibrium theory (GET) is the economics of yesterday. While GET had mirrored most analytic advances in economic theory before Debreu, after Debreu most theoretical innovations came as alternatives to GET (from game theory to complexity theory)". Historian of economic thought Roger Backhouse writes that "[i]n the 1940s and 1950s general-equilibrium theory [...] became seen as the central theoretical framework around which economics was based" (Backhouse 2002: 254) and that by the "[...] early 1960s, confidence in general-equilibrium theory, and with it economics as a whole, as at its height, with Debreu's *Theory of Value* being widely seen as providing a rigorous, axiomatic framework at the centre of the discipline" (Backhouse 2002: 261), but "[...] there were problems that could not be tackled within the Arrow-Debreu framework. These include money (attempts were made to develop a general-equilibrium theory of money, but they failed), information, and imperfect competition. In order to tackle such problems, economists were forced to use less general models, often dealing only with a specific part of the economy or with a particular problem. The search for ever more general models of general competitive equilibrium, that culminated in *Theory of Value*, was over". (Backhouse 2002: 262). One set of particularly problematic results for general equilibrium are the Sonnenschein-Mantel-Debreu (SMD) theorems. "In part because of a conviction that progress could not be made in general equilibrium theory, there was a substantial redirection in economic theory. As the results in SMD theory became well known, for example through Wayne Shafer and Hugo Sonnenschein's survey (1982), economists began to question the centrality of general equilibrium theory and put forward alternatives to it. Thus in the ten years following the Shafer-Sonnenschein survey, we find a number of new directions in economic theory". (Rizvi 2006: 230).

⁸²Gale (2000: 38-9) describes partial equilibrium in the following terms, "[...] partial equilibrium analysis, that amalgam of handy short cuts that allows economists to isolate particular phenomena and study them on the back of a virtual envelope, ignoring the fact that an economy is a complex system in which "everything affects and is affected by everything else". It may not be pure, but it is very practical". He describes the "elegant theory" of GE as a theory in which "[...] the interaction of individual agents and individual markets throughout the economy are aggregated to provide a precise account of the equilibrium of the entire economy. It may not be practical, but it is very pure". (Gale 2000: 39).

⁸³This is not to say there has been no work at all on GE within these areas. For an example of work on a GE approach to firms see Zame (2007). For a discussion of the Arrow-Debreu model when faced with moral hazard and adverse selection see Guesnerie (1992). For a look at effects to provide strategic foundations for GE see Gale (2000). On the GE approach to tax and to international trade see, for example, Shoven and Whalley (1984), Creedy (1997), Jones (2011) and Woodland (2011). There has also been much work on computable general equilibrium analysis, see Boehringer, Rutherford and Wiegard (2003) and Sue Wing (2004) for overviews.

⁸⁴Arrow (1971) discusses the pre-1970 GE approach to the firm.

on necessarily partial models that take into account the full complexity of strategic interactions between privately informed agents in well-defined institutional settings”.

In the previous section it was noted that the current literature on the theory of the firm can be classified into two general groups based on which of two of the standard assumptions of GE theory, namely symmetric information and complete contracts, is violated when modelling the firm. In contract theoretic terms the former gives rise to principal-agent type models and the latter to incomplete contract type theories.

The necessity of having to violate basic assumptions of GE theory so that we can model the firm, suggests that as it stands GE can not deal easily with firms, or other important economic institutions. Bernard Salanié has noted that,

“[. . .] the organization of the many institutions that govern economic relationships is entirely absent from these [GE] models. This is particularly striking in the case of firms, which are modeled as a production set. This makes the very existence of firms difficult to justify in the context of general equilibrium models, since all interactions are expected to take place through the price system in these models”. (Salanié 2005: 1).

This would suggest that creating GE models that can account for information asymmetries, contractual incompleteness, strategic interaction and the existence of institutions is needed if GE theory is to be able to deal with the issues – including those related to the firm – that microeconomists wish to tackle.

References

- Aghion, Philippe and Richard Holden (2011). 'Incomplete Contracts and the Theory of the Firm: What Have We Learned Over the Past 25 Years?', *Journal of Economic Perspectives*, 25(2) Spring: 181-97.
- Alchian, Armen (1984). 'Specificity, specialization, and coalitions', *Journal of Institutional and Theoretical Economics*, 140(1) March: 34-49.
- Alchian, Armen and Harold Demsetz (1972). 'Production, Information Costs, and Economic Organization', *American Economic Review*, 62(5) December: 777-95.
- Allen, Douglas and Dean Lueck (2002). *The Nature of the Farm: Contracts, Risk, and Organization in Agriculture*, Cambridge Mass.: MIT Press.
- Allen, G. C. (1929). *The Industrial Development of Birmingham and the Black Country, 1860-1927*, London: George Allen & Unwin Ltd.
- Argote, Linda and Henrich R. Greve (2007). 'A Behavioral Theory of the Firm—40 Years and Counting: Introduction and Impact', *Organization Science*, 18(3) May-June: 337-49.
- Arrow, Kenneth J. (1971). 'The Firm in General Equilibrium Theory'. In R. Marris and A. Wood, (eds.), *The Corporate Economy: Growth, Competition, and Innovative Potential* (pp. 68-110), London: Macmillan.
- Augier, Mie and James G. March (2008). 'A retrospective look at A Behavioral Theory of the Firm', *Journal of Economic Behavior & Organization*, 66(1) April: 1-6.
- Backhouse, Roger E. (2002). *The Ordinary Business of Life: A History of Economics from the Ancient World to the Twenty-First Century*, Princeton: Princeton University Press.
- Backhouse, Roger E. (2003). 'The Stabilization of Price Theory, 1920-1955'. In Warren J. Samuels, Jeff E. Biddle and John B. Davis (eds.), *A Companion to the History of Economic Thought* (pp. 308-24), Oxford: Blackwell Publishing Ltd.
- Baker, Dean and Mark Weisbrot (1994). 'The Logic of Contested Exchange', *Journal of Economic Issues*, 28(4) December: 1091-114.
- Baker, George, Robert Gibbons and Kevin J. Murphy (2002). 'Relational Contracts and the Theory of the Firm', *Quarterly Journal of Economics*, 117(1) February: 39-84.
- Barba Navaretti, Giorgio and Anthony J. Venables with Frank G. Barry, Karolina Ekholm, Anna M. Falzoni, Jan I. Haaland, Karen Helene Midelfart, and Alessandro Turrini (2004). *Multinational Firms in the World Economy*, Princeton: Princeton University Press.
- Barker, Graeme (2006). *The Agricultural Revolution in Prehistory: Why did Forages Become Farmers?*, Oxford: Oxford University Press.
- Barney, Jay (1991). 'Firm Resources and Sustained Competitive Advantage', *Journal of Management*, 17(1) March: 99-120.
- Baran, Paul A. and Paul M. Sweezy (1966). *Monopoly Capital: An Essay on the American Economic and Social Order*, New York: Monthly Review Press.
- Barzel, Yoram (1987). 'Knight's "Moral Hazard" Theory of Organization', *Economic Inquiry*, 25(1) January: 117-20.
- Barzel, Yoram (1997). *Economic Analysis of Property Rights*, 2nd edn., Cambridge: Cambridge University Press.

- Baumol, William J. (1962). 'On the Theory of Expansion of the Firm', *American Economic Review*, 52(5) December: 1078-87.
- Baumol, William J. (1967). *Business Behavior, Value and Growth*, rev. ed., New York: Macmillan.
- Berle, Adolf Augustus Jr. and Gardiner C. Means (1932). *The Modern Corporation and Private Property*, New York: Macmillan.
- Berry, Arthur (1891). 'The Pure Theory of Distribution'. In *Report of the Sixtieth Meeting of the British Association for the Advancement of Science Held at Leeds September 1890* (pp. 923-4), London: John Murray.
- Blaug, Mark (1958). 'The Classical Economists and the Factory Acts-A Re-Examination', *The Quarterly Journal of Economics*, 72(2) May: 211-26.
- Blaug, Mark (1996). *Economic Theory in Retrospect*, 5th edn., Cambridge: Cambridge University Press.
- Böhringer, Christoph, Thomas F. Rutherford and Wolfgang Wiegand (2003). 'Computable General Equilibrium Analysis: Opening a Black Box', ZEW Discussion Paper No. 03-56, Mannheim, Germany.
- Bolton, Patrick and Mathias Dewatripont (1994). 'The Firm as a Communication Network', *The Quarterly Journal of Economics*, 109(4) November: 809-40.
- Boudreaux, Donald J. and Randell G. Holcombe (1989). 'The Coasian and Knightian Theories of the Firm', *Managerial and Decision Economics*, 10(2) June: 147-54.
- Borland, Jeff and Gerald Garvey (1994). 'Recent Developments in the Theory of the Firm', *Australian Economic Review*, issue 105 January-March: 60-82.
- Bös, Dieter (1991). *Privatization: A Theoretical Treatment*, Oxford: Oxford University Press.
- Boulding, Kenneth E. (1942). 'The Theory of the Firm in the Last Ten Years', *American Economic Review*, 32(4) December: 791-802.
- Boulding, Kenneth E. (1960). 'The present position of the theory of the firm'. In Kenneth E. Boulding and W. Allen Spivey (eds.), *Linear programming and the theory of the firm* (pp. 1-17), New York: The Macmillan Company.
- Boycko, Maxim, Andrei Shleifer and Robert W. Vishny (1996). 'A Theory of Privatisation', *The Economic Journal*, 106 no. 435 March: 309-19.
- Butler, Eamonn (2007). *Adam Smith - A Primer*, London: Institute of Economic Affairs.
- Bylund, Per (2011). 'Division of Labor and the Firm: An Austrian Attempt at Explaining the Firm in the Market', *Quarterly Journal of Austrian Economics*, 14(2): 188-215.
- Cannan, Edwin (1903). *A History of the Theories of Production and Distribution in English Political Economy from 1776 TO 1848*, 2nd edn., London: P. S. King & Son.
- Carlos, Ann M. and Stephen Nicholas (1996). 'Theory and History: Seventeenth-Century Joint-Stock Chartered Trading Companies', *The Journal of Economic History*, 56(4) December: 916-24.
- Cawston, George and A. H. Keane (1896). *The Early Chartered Companies (A.D. 1296-1858)*, London: Edward Arnold.
- Chamberlin, Edward H. (1933). *The Theory of Monopolistic Competition*, Cambridge, Mass.: Harvard University Press.

- Chamberlin, Edward H. (1937). 'Monopolistic or Imperfect Competition?', *Quarterly Journal of Economics*, 51(4) August: 557-80.
- Chandler, Alfred D., Jr. (1962). *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*, Cambridge, Mass: MIT Press
- Chandler, Alfred D., Jr. (1977). *The Visible Hand: The Managerial Revolution in American Business*, Cambridge Mass.: Harvard University Press.
- Chandler, Alfred D., Jr. (1990). *Scale and Scope: The Dynamics of Industrial Capitalism*, Cambridge, Mass.: Harvard University Press, Belknap Press.
- Chapman, Stanley D. (1974). 'The Textile Factory before Arkwright: A Typology of Factory Development', *Business History Review*, 48(4) Winter: 451-78.
- Cheung, Steven N.S. (1983). 'The Contractual Nature of the Firm', *Journal of Law and Economics*, 26(1) April: 1-21.
- Clapham, J. H. (1922). 'Of Empty Economic Boxes', *The Economic Journal*, 32 no. 127 September: 305-4.
- Coad, Alexander (2007). 'Firm Growth: A Survey', Evolutionary Economics Group, Max Planck Institute of Economics, Jena, Papers on Economics and Evolution # 0703, May 14.
- Coad, Alexander (2009). *The Growth of Firms: A Survey of Theories and Empirical Evidence*, Cheltenham, UK: Edward Elgar.
- Coase, Ronald Harry (1937). 'The Nature of the Firm', *Economica*, n.s. 4 no. 16 November: 386-405.
- Coase, Ronald Harry (1955). 'Comment' (on Heflebower (1955)). In Universities-National Bureau, *Business Concentration and Price Policy* (pp. 392-4), Princeton: Princeton University Press.
- Coase, Ronald Harry (1960). 'The Problem of Social Cost', *Journal of Law and Economics*, 3 October: 1-44.
- Coase, Ronald Harry (1972). 'Industrial Organization: A Proposal for Research'. In Victor R. Fuchs (ed.), *Economic Research: Retrospect and Prospect* (vol. 3: Policy Issues and Research Opportunities in Industrial Organization, pp. 59 - 73), Cambridge Mass.: National Bureau of Economic Research.
- Coase, Ronald Harry (1988a). 'The Nature of the Firm: Meaning', *Journal of Law, Economics & Organization*, 4(1) Spring: 19-32.
- Coase, Ronald Harry (1988b). 'The Nature of the Firm: Influence', *Journal of Law, Economics & Organization*, 4(1) Spring: 33-47.
- Colander, David, Richard Holt and J. Barkley Rosser Jr. (2004). 'The changing face of mainstream economics', *Review of Political Economy*, 16(4): 485-99.
- Colander, David, Richard Holt and J. Barkley Rosser Jr. (2005). *The changing face of economics: conversations with cutting edge economists*, Ann Arbor: University of Michigan Press.
- Conner, K. R. (1991). 'A Historical Comparison of Resource-Based Theory and Five Schools of Thought Within Industrial Organization: Do We Have a New Theory of the Firm?', *Journal of Management*, 17: 121-54.
- Conner, K. R. and C. K. Prahalad (1996). 'A Resource-Based Theory of the Firm: Knowledge versus Opportunism', *Organization Science*, 7(5) September-October: 477-501.

- Copp, Stephen F. (2008). 'Limited Liability and Freedom'. In Stephen F. Copp (ed.), *The Legal Foundations of Free Markets* (pp. 158-90), London: Institute of Economic Affairs.
- Creedy, John (1997). 'Taxation in General Equilibrium: an Introduction', *Bulletin of Economic Research*, 49(3) July: 177-203.
- Cyert, Richard M. and Charles L. Hedrick (1972). 'Theory of the Firm: Past, Present, and Future; An Interpretation', *Journal of Economic Literature*, 10(2) June: 398-412.
- Cyert, Richard M. and James G. March (1963). *A Behavioral Theory of the Firm*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Debreu, Gerard (1959). *Theory of Value*, New York: Wiley.
- Demsetz, Harold (1982). *Economic, Legal, and Political Dimensions of Competition*, Amsterdam: North-Holland Publishing Company.
- Demsetz, Harold (1988a). 'The Theory of the Firm Revisited', *Journal of Law, Economics, & Organization*, 4(1) Spring: 141-61.
- Demsetz, Harold (1988b). 'Profit as a Functional Return: Reconsidering Knight's Views'. In Harold Demsetz, *Ownership, Control, and the Firm: The Organization of Economic Activity*, (vol. I, pp. 236-47), Oxford: Basil Blackwell.
- Demsetz, Harold (1995). *The Economics of the Business Firm: Seven Critical Commentaries*, Cambridge: Cambridge University Press.
- Demsetz, Harold (1996). 'The core disagreement between Pigou, the profession, and Coase in the analyses of the externality question', *European Journal of Political Economy*, 12(4) December: 565-79.
- Demsetz, Harold (1997). 'The Firm in Economic Theory: A Quiet Revolution', *American Economic Review*, 87(2) May: 426-29.
- Dequech, David (2007). 'Neoclassical, mainstream, orthodox, and heterodox economics', *Journal of Post Keynesian Economics*, 30(2): 279-302.
- De Scitovszky, T. (1943). 'A Note on Profit Maximisation and its Implications', *The Review of Economic Studies*, 11(1) Winter: 57-60.
- Dulbecco, Philippe and Pierre Garrouste (1999). 'Towards an Austrian Theory of the Firm', *Review of Austrian Economics*, 12(1): 43-64.
- Düppe, Till (2010). 'Debreu's apologies for mathematical economics after 1983', *Erasmus Journal for Philosophy and Economics*, 3(1) Spring: 1-32.
- Edgeworth, Francis Y. (1889). 'On the Application of Mathematics to Political Economy', *Journal of the Royal Statistical Society*, 52(4) December: 538-76.
- Ekelund, Robert Jr. and Robert D. Tollison (1997). *Politicized Economies: Monarchy, Monopoly, and Mercantilism*, College Station: Texas A&M University Press.
- Evensky, J. (2005). *Adam Smith's Moral Philosophy: A Historical and Contemporary Perspective on Markets, Law, Ethics, and Culture*, Cambridge: Cambridge University Press.
- Fama, Eugene F. (1980). 'Agency Problems and the Theory of the Firm', *Journal of Political Economy*, 88(2) April: 288-307.
- Fehr, Ernst, Oliver D. Hart and Christian Zehnder (2009). 'Contracts, Reference Points, and Competition – Behavior Effects of the Fundamental Transformation', *Journal of the European Economic Association*, 7(2) April-May: 561-72.

- Fehr, Ernst, Oliver D. Hart and Christian Zehnder (2011). 'Contracts as Reference Points – Experimental Evidence', *American Economic Review*, 101(2) May: 493-525.
- Foss, Kirsten and Nicolai J. Foss (2006). 'Entrepreneurship, Transaction Costs, and Resource Attributes', Center for Strategic Management and Globalization, Copenhagen Business School, SMG Working Paper No. 7/2006, April.
- Foss, Nicolai J. (1993). 'More on Knight and the Theory of the Firm', *Managerial and Decision Economics*, 14(3) May-June: 269-76.
- Foss, Nicolai J. (1994). 'The Theory of the Firm: The Austrians as Precursors and Critics of Contemporary Theory', *Review of Austrian Economics*, 7: 31-64.
- Foss, Nicolai J. (1996a). 'Knowledge-Based Approaches to the Theory of the Firm: Some Critical Comments', *Organization Science*, 7(5) September-October: 470-6.
- Foss, Nicolai J. (1996b). 'More Critical Comments on Knowledge-Based Theories of the Firm', *Organization Science*, 7(5) September-October: 519-23.
- Foss, Nicolai J. (1996c). 'Harald B. Malmgren's Analysis of the Firm: lessons for modern theorists?', *Review of Political Economy*, 8(4): 349-66.
- Foss, Nicolai J. (1997b). 'Austrian Insights and the Theory of the Firm', *Advances in Austrian Economics*, 4: 175-98.
- Foss, Nicolai J. (2000). 'The Theory of the Firm: An Introduction to Themes and Contributions'. In Nicolai Foss (ed.), *The Theory of the Firm: Critical Perspectives on Business and Management* (pp. xv-lxi), London: Routledge.
- Foss, Nicolai J. and Peter G. Klein (2006). 'The Emergence of the Modern Theory of the Firm', Center for Strategic Management and Globalization, Copenhagen Business School, SMG Working Paper 1/2006, January.
- Foss, Nicolai J. and Peter G. Klein (2008). 'The Theory of the Firm and Its Critics: a Stocktaking and an Assessment'. In Eric Brousseau and Jean-Michel Glachant (eds.), *New Institutional Economics: A Guidebook* (pp. 425-42), Cambridge: Cambridge University Press.
- Foss, Nicolai J. and Peter G. Klein (2009). 'Austrian Economics and the Transaction Cost Approach to the Firm', *Libertarian Papers*, 1 article number 39.
- Foss, Nicolai J. and Peter G. Klein (2010). 'Austrian Economics and the Theory of the Firm'. In Peter G. Klein and Michael E. Sykuta (eds.), *The Elgar Companion to Transaction Cost Economics* (pp. 281-96), Cheltenham U.K.: Edward Elgar Publishing Ltd.
- Foss, Nicolai J., Peter G. Klein and Per L. Bylund (2011). 'Entrepreneurship and the Economics of the Firm'. In Daniel Hjorth (ed.), *Handbook of Organisational Entrepreneurship*, Cheltenham U.K.: Edward Elgar, Forthcoming, available from SSRN: <http://ssrn.com/abstract=1747710>
- Foss, Nicolai J., Henrik Lando and Steen Thomsen (2000). 'The Theory of the Firm'. In Boudewijn Bouckaert and Gerrit De Geest (eds.), *Encyclopedia of Law and Economics* (vol. III, pp. 631-58), Cheltenham U.K.: Edward Elgar Publishing Ltd.
- Foss, Nicolai J. and Nils Stieglitz (2010). 'Modern Resource-Based Theory(ies)', Center for Strategic Management and Globalization, Copenhagen Business School, SMG Working Paper 7/2010, September.
- Galbraith, John Kenneth (1963). *American Capitalism: The Concept of Countervailing Power*, New York: Penguin Books.

- Galbraith, John Kenneth (1969). *The New Industrial State*, New York: Penguin Books.
- Gale, Douglas (2000). *Strategic Foundations of General Equilibrium: Dynamic Matching and Bargaining Games* (Churchill Lectures in Economic Theory), Cambridge: Cambridge University Press.
- Garrouste, Pierre (2004). 'The new property rights theory of the firm'. In E. Colombatto (ed.), *The Edward Elgar Companion to the Economics of Property Rights* (pp. 370-82), Cheltenham U.K.: Edward Elgar Publishing Ltd.
- Garrouste, Pierre and Stéphane Saussier (2008). 'The Theories of the Firm'. In Eric Brousseau and Jean-Michel Glachant (eds.), *New Institutional Economics: A Guidebook* (pp. 23-36), Cambridge: Cambridge University Press.
- Gibbons, Robert (2005). 'Four formal(izable) theories of the firm?', *Journal of Economic Behavior & Organization*, 58(2): 200-45.
- Gibbons, Robert (2010). 'Transaction-Cost Economics: Past, Present, and Future?', *Scandinavian Journal of Economics*, 112(2): 263-88.
- Griffiths, Percival (1974). *A Licence to Trade: A History of the English Chartered Companies*, London: Ernest Benn Limited.
- Grossman, Sanford J. and Oliver D. Hart (1986). 'The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration', *Journal of Political Economy*, 94(4): 691-719.
- Grossman, Sanford J. and Oliver D. Hart (1987). 'Vertical Integration and the Distribution of Property Rights'. In Assaf Razin and Efraim Sadka (eds.), *Economic Policy in Theory and Practice* (pp. 504-48), London: Macmillan Press.
- Guesnerie, Roger (1992). 'The Arrow-Debreu Paradigm Faced with Modern Theories of Contracting: a Discussion of Selected Issues Involving Information and Time'. In Lars Werin and Hans Wijkander (eds.), *Contract Economics* (pp. 12-41), Oxford: Basil Blackwell Ltd.
- Hagendorf, Klaus (2009). 'Labour Values and the Theory of the Firm, Part I: The Competitive Firm', October 15. Available at SSRN: <http://ssrn.com/abstract=1489383>.
- Hall, R. L. and C. J. Hitch (1939). 'Price Theory and Business Behaviour', *Oxford Economic Papers*, 1939(2) May: 12-45.
- Hansmann, Henry (1996). *The Ownership of Enterprise*, Cambridge, Mass.: Harvard University Press.
- Hansmann, Henry and Reinier Kraakman (2000a). 'The Essential Role of Organizational Law', *Yale Law Journal*, 110(3) December: 387-440.
- Hansmann, Henry and Reinier Kraakman (2000b). 'Organizational Law as Asset Partitioning', *European Economic Review*, 44: 807-17.
- Hansmann, Henry, Reinier Kraakman and Richard Squire (2005). 'The New Business Entities in Evolutionary Perspective', *University of Illinois Law Review*, 2005(1): 5-14.
- Hansmann, Henry, Reinier Kraakman and Richard Squire (2006). 'Law and the Rise of the Firm', *Harvard Law Review*, 119(5) March: 1333-403.
- Hart, Neil (2003). 'Marshall's Dilemma: Equilibrium versus Evolution', *Journal of Economic Issues*, 37(4) December: 1139-60.
- Hart, Oliver D. (1989). 'An Economist's Perspective on the Theory of the Firm', *Columbia Law Review*, 89(7) November: 1757-74.

- Hart, Oliver D. (1995). *Firms, Contracts, and Financial Structure*, Oxford: Oxford University Press.
- Hart, Oliver D. (2003). 'Incomplete Contracts and Public Ownership: Remarks, and an Application to Public-Private Partnerships', *The Economic Journal*, 113 No. 486 Conference Papers March: C69-C76.
- Hart, Oliver D. (2008). 'Economica Coase Lecture: Reference Points and the Theory of the Firm', *Economica*, 75(299) August: 404-11.
- Hart, Oliver D. (2009). 'Hold-up, Asset Ownership, and Reference Points', *Quarterly Journal of Economics*, 124(1) February: 267-300.
- Hart, Oliver D. (2011a). 'Noncontractible Investments and Reference Points', *National Bureau of Economic Research*, NBER Working Papers: 16929.
- Hart, Oliver D. (2011b). 'Thinking about the Firm: A Review of Daniel Spulber's The Theory of the Firm', *Journal of Economic Literature*, 49(1) March: 101-13.
- Hart, Oliver D. and Bengt Holmström (2010). 'A Theory of Firm Scope', *Quarterly Journal of Economics*, 125 (2) May: 483-513 .
- Hart, Oliver D. and John Moore (1990). 'Property Rights and the Nature of the Firm', *Journal of Political Economy*, 98(6): 1119-58.
- Hart, Oliver D. and John Moore (2007). 'Incomplete Contracts and Ownership: Some New Thoughts', *American Economic Review*, 97(2) May: 182-6.
- Hart, Oliver D. and John Moore (2008). 'Contracts as Reference Points', *Quarterly Journal of Economics*, 123(1) February: 1-48.
- Hart, Oliver D., Andrei Shleifer and Robert W. Vishny (1997). 'The Proper Scope of Government: Theory and an Application to Prisons', *Quarterly Journal of Economics*, 112(4) November: 1127-61.
- Heckscher, Eli F. (1934). *Mercantilism*, 2 vols., London: George Allen & Unwin.
- Heflebower, Richard B. (1955). 'Full Costs, Cost Changes, and Prices'. In Universities-National Bureau, *Business Concentration and Price Policy* (pp. 359-92), Princeton: Princeton University Press.
- Hicks, John R. (1946). *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*, 2nd edn., Oxford: Clarendon Press.
- Hickson, Charles R. and John D. Turner (2006). 'Corporation or Limited Liability Company'. In John J. McCusker (ed.), *History of World Trade since 1450* (vol. 1 A-K, pp. 163-66), New York: Macmillan Reference USA.
- Holmström, Bengt (1982). 'Moral Hazard in Teams', *Bell Journal of Economics*, 13(2) Autumn: 324-40.
- Holmström, Bengt and Paul Milgrom (1991). 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design', *Journal of Law, Economics, & Organization*, 7(Special Issue): 24-52.
- Holmström, Bengt and Paul Milgrom (1994). 'The Firm as an Incentive System', *American Economic Review*, 84(4) September: 972-91.
- Holmström, Bengt and John Roberts (1998). 'Boundaries of the Firm Revisited', *Journal of Economic Perspectives*, 12(4) Fall: 73-94.

- Holmström, Bengt and Jean Tirole (1989). 'The Theory of the Firm'. In Richard Schmalensee and Robert D. Willig (eds.), *Handbook of Industrial Organization* (vol. I, pp. 61-133), Amsterdam: North-Holland.
- Hoppe, Eva I. and Patrick W. Schmitz (2011). 'Can contracts solve the hold-up problem? Experimental evidence', *Games and Economic Behavior*, 73(1) September: 186-99.
- Hubbard, Thomas N. (2008). 'Viewpoint: Empirical Research on Firms' Boundaries', *Canadian Journal of Economics*, 41(2) May: 341-59.
- Humphrey, Thomas M. (1997). 'Algebraic Production Functions and Their Uses Before Cobb-Douglas', *Federal Reserve Bank of Richmond Economic Quarterly*, 83(1) Winter: 51-83.
- Ioannides, Stavros (1999). 'Towards an Austrian Perspective on the Firm', *Review of Austrian Economics*, 11(1-2): 77-97.
- Jacobides, M. and S. Winter (2005). The Co-evolution of Capabilities and Transaction Costs: Explaining the Institutional Structure of Production. *Strategic Management Journal*, 26(5) May: 395-413.
- Jankovic, Ivan (2010). 'Firm as a Nexus of Markets', *Journal des Economistes et des Etudes Humaines*, 16(1): Article 3. Available at: <http://www.bepress.com/jeeh/vol16/iss1/art3>
- Jensen, Michael C. and William H. Meckling (1976). 'Theory of the Firm: Managerial Behaviour, Agency Costs, and Ownership Structure', *Journal of Financial Economics*, 3(4) October: 305-60.
- Johnson, William E. (1891). 'Exchange and Distribution', *Cambridge Economic Club*. Reprinted, with brief commentary, in William J. Baumol and Stephen N. Goldfeld (eds.), *Precursors in Mathematical Economics: An Anthology* (pp. 316-20), London: The London School of Economics and Political Science, 1968 and in A. C. Darnell (ed.), *Early Mathematical Economics* (vol. 2, pp. 67-71) London: Pickering & Chatto, 1991.
- Jones, Ronald W. (2011). 'General Equilibrium Theory and Competitive Trade Models', paper given at the European Trade Study Group (ETSG), Thirteenth Annual Conference, Copenhagen Business School, University of Copenhagen, 8-10 September.
- Joskow, Paul L. (1988). 'Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence', *Journal of Law, Economics, & Organization*, 4(1) Spring: 95-117.
- Jones, Eliot (1921). *The Trust Problem in the United States*, New York: Macmillan Co.
- Jones, S. R. H. and Simon P. Ville (1996a). 'Efficient Transactors or Rent-Seeking Monopolists? The Rationale for Early Chartered Trading Companies', *The Journal of Economic History*, 56(4) December: 898-915.
- Jones, S. R. H. and Simon P. Ville (1996b). 'Theory and Evidence: Understanding Chartered Trading Companies', *The Journal of Economic History*, 56(4) December: 925-6.
- Kennedy, Gavin (2005). *Adam Smith's Lost Legacy*, Houndmills Basingstoke Hampshire: Palgrave Macmillan.
- Kennedy, Gavin (2010). *Adam Smith: A Moral Philosopher and His Political Economy*, 2nd edn., Houndmills Basingstoke Hampshire: Palgrave Macmillan.
- Khachatryan, Arman (2003). 'Theory of the firm: a critical survey', Siena Memos and Papers in Law and Economics (SIMPLE) 4/03, March.

- Khanna, Vikramaditya S. (2005). 'The Economic History of Organizational Entities in Ancient India', University of Michigan Law School Program in Law & Economics Working Paper No. 14, updated version 2/2006.
- Klein, Benjamin, Robert G. Crawford and Armen A. Alchian (1978). 'Vertical Integration, Appropriable Rents, and the Competitive Contracting Process', *Journal of Law and Economics*, 21(2) October: 297-326.
- Klein, Peter (1996). 'Economic Calculation and the Limits of Organization', *The Review of Austrian Economics*, 9(2): 3-28.
- Klein, Peter (2005). 'The Make-or-Buy Decision: Lessons from Empirical Studies'. In Claude Menard and Mary M. Shirley (eds.), *Handbook of New Institutional Economics* (pp. 435-64), Dordrecht and New York: Springer.
- Klein, Peter G. (2010). *The Capitalist and The Entrepreneur: Essays on Organizations and Markets*, Auburn: Ludwig von Mises Institute.
- Knight, Frank H. (1921a). 'Cost of Production and Price over Long and Short Periods', *Journal of Political Economy*, 29 April: 304-35.
- Knight, Frank H. (1921b). *Risk, Uncertainty and Profit*, Boston: Houghton Mifflin Company.
- Kogut, B. and U. Zander (1992). 'Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology', *Organization Science*, 3(3) August: 383-97.
- Kogut, B. and U. Zander (1996). 'What Firms Do? Coordination, Identity, and Learning', *Organization Science*, 7(5) September-October: 502-18.
- Koutsoyiannis, A. (1979). *Modern Microeconomics*, 2nd edn., London: The Macmillan Press Ltd.
- Kuran, Timur (2005). 'The Absence of the Corporation in Islamic Law: Origins and Persistence', *American Journal of Comparative Law*, 53(4) Fall: 785-834.
- Laffone, Jean-Jacques and Jean Tirole (1991). 'Privatization and Incentives', *Journal of Law, Economics, & Organization*, 7 (Special Issue) [Papers from the Conference on the New Science of Organization, January 1991]: 84-105.
- Lafontaine, Francine and Margaret Slade (2007). 'Vertical Integration and Firm Boundaries: The Evidence', *Journal of Economic Literature*, 45(3) September: 629-85.
- Lal, Deepak (2006). 'The Contemporary Relevance of Heckscher's Mercantilism'. In Ronald Findlay, Rolf G. H. Henriksson, Håkan Lindgren and Mats Lundahl (eds.), *Eli Heckscher, International Trade, and Economic History* (pp. 305-19), Cambridge Mass.: MIT Press.
- Langlois, Richard N. (1981). 'Why are there firms?', C. V. Starr Center for Applied Economics, New York University, D.P. 81-30, March.
- Langlois, Richard N. (forthcoming). 'The Austrian Theory of the Firm: Retrospect and Prospect', *Review of Austrian Economics*.
- Langlois, Richard N. and Metin M. Cosgel (1993). 'Frank Knight on Risk, Uncertainty, and the Firm: A New Interpretation', *Economic Inquiry*, 31(3) July: 435-65.
- Lee, F. S. (1984). 'The Marginalist Controversy and the Demise of Full-Cost Pricing', *Journal of Economic Issues*, 18(4) December: 1107-32.
- Leibenstein, Harvey (1966). 'Allocative Efficiency vs. X-Efficiency', *American Economic Review*, 56 (3) June: 392-415.

- Lester, Richard A. (1946). 'Shortcomings of Marginal Analysis for Wage-Employment Problems', *American Economic Review*, 36(1) March: 63-82.
- Lewin, Peter and Steven E. Phelan (2000). 'An Austrian Theory of the Firm', *Review of Austrian Economics*, 13(1): 59-79.
- Liebhafsky, H. H. (1955). 'A Curious Case of Neglect: Marshall's Industry and Trade', *Canadian Journal of Economics and Political Science*, 21(3) August: 339-53.
- Lockett, Andy, Rory P. O'Shea and Mike Wright (2008). 'The Development of the Resource-based View: Reflections from Birger Wernerfelt', *Organization Studies*, 29(8-9) August: 1125-41.
- Macgregor, D. H. (1906). *Industrial Combination*, London: G. Bell & Sons.
- Machlup, Fritz (1946). 'Marginal Analysis and Empirical Research', *American Economic Review*, 36(4) September: 519-54.
- Machlup, Fritz (1967). 'Theories of the Firm: Marginalist, Behavioral, Managerial', *American Economic Review*, 57(1) March: 1-33.
- Maekawa, Kazuya (1980). 'Female Weavers and Their Children in Lagash: Pre-Sargonic and Ur III', *Acta Sumerologica*, 21: 81-125.
- Magnusson, Lars G. (1994). *Mercantilism: The Shaping of Economic Language*, London: Routledge.
- Magnusson, Lars G. (2003). 'Mercantilism'. In Warren J. Samuels, Jeff E. Biddle and John B. Davis (eds.), *A Companion to the History of Economic Thought* (pp. 46-60), Oxford: Blackwell Publishing Ltd.
- Mahoney, Joseph T. (2005). *Economic Foundations of Strategy*, London: Sage Publication.
- Malmgren, H. B. (1961). 'Information, Expectations and the Theory of the Firm', *The Quarterly Journal of Economics*, 75(3) August: 399-421.
- Marchionatti, Roberto (2003). 'On the Methodological Foundations of Modern Microeconomics: Frank Knight and the "Cost Controversy" in the 1920s', *History of Political Economy*, 35(1): 49-75.
- Marris, Robin L. (1964). *The Economic Theory of Managerial Capitalism*, London: Macmillan.
- Marshall, Alfred (1920). *Industry and Trade: A Study of Industrial Technique and Business Organization; and of Their Influences on the Conditions of Various Classes and Nations*, 2 vols., Honolulu: University Press of the Pacific, 2003.
- Mas-Colell, A., M. D. Whinton and J. R. Green (1995). *Microeconomic Theory*, New York: Oxford University Press.
- Maskin, Eric (2002). 'On indescribable contingencies and incomplete contracts', *European Economic Review*, 46(4-5) May: 725-33.
- Maskin, Eric and Jean Tirole (1999). 'Unforeseen Contingencies and Incomplete Contracts', *Review of Economic Studies*, 66(1) January: 83-114.
- McAfee, R. Preston and John McMillan (1991). 'Optimal Contracts for Teams', *International Economic Review*, 32(3) August: 561-77.
- McManus, John C. (1975). 'The Costs of Alternative Economic Organizations', *Canadian Journal of Economics*, 8(3) August: 334-50.
- McNulty, Paul J. (1984). 'On the Nature and Theory of Economic Organization: the Role of the Firm Reconsidered', *History of Political Economy*, 16(2) Summer: 233-53.

- Menard, Claude (2005). 'A New Institutional Approach to Organization'. In Claude Menard and Mary Shirley (eds.), *Handbook of New Institutional Economics* (pp. 281-318), Dordrecht and New York: Springer.
- Micklethwait, John and Adrian Wooldridge (2003). *The Company: A Short History of a Revolutionary Idea*, New York: The Modern Library.
- Milgrom, Paul and John Roberts (1988). 'Economic Theories of the Firm: Past, Present, and Future', *Canadian Journal of Economics*, 21(3) August: 444-58.
- Mokyr, Joel (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton: Princeton University Press.
- Mokyr, Joel (2009). *The Enlightened Economy: An Economic History of Britain 1700-1850*, New Haven: Yale University Press.
- Mongin, Philippe (1992). 'The Full-Cost Controversy of the 1940s and 1950s: A Methodological Assessment', *History of Political Economy*, 24(2) Summer: 311-56.
- Mongin, Philippe (1998). 'The Marginalist Controversy'. In John B. Davis, D. Wade Hands and Uskali Mäki (eds.), *Handbook of Economic Methodology* (pp. 277-81), Cheltenham, UK: Edward Elgar.
- Moore, John (1992). 'The Firm as a Collection of Assets', *European Economic Review*, 36(2-3) April: 493-507.
- Moss, Scott (1984a). 'The History of the Theory of the Firm from Marshall to Robinson and Chamberlin: the Source of Positivism in Economics', *Economica*, n.s. 51 no. 203 August: 307-18.
- Moss, Scott (1984b). 'O'Brien's 'The Evolution of the Theory of the Firm': a Discussion'. In Frank H. Stephen (ed.), *Firms, Organization and Labour: Approaches to the Economics of Work Organization* (pp. 63-8), London: Macmillan Press.
- Müller, Alberto (2009). 'Teoría de la firma, teoría del mercado y teoría económica: Una reflexión', *Desarrollo Económico*, 49 issue 194 July-September: 335-50.
- Nelson, Richard R. and Sidney G. Winter (1982). *An Evolutionary Theory of Economic Change*, Cambridge, Mass.: The Belknap Press.
- O'Brien, D. P. (1984). 'The Evolution of the Theory of the Firm'. In Frank H. Stephen (ed.), *Firms, Organization and Labour: Approaches to the Economics of Work Organization* (pp. 25-62), London: Macmillan Press.
- Ofek, Haim (2001). *Second Nature: Economic Origins of Human Evolution*, Cambridge: Cambridge University Press.
- Otteson, J. R. (2002). *Adam Smith's Market Place of Life*, Cambridge: Cambridge University Press.
- Otteson, J. R. (2011). *Adam Smith* (Major Conservative and Libertarian Thinkers Volume 16, Series Editor: John Meadowcroft), New York: Continuum.
- Papandreou, Andreas G. (1952). 'Some Basic Problems in the Theory of the Firm'. In Bernard F. Haley (ed.), *A Survey of Contemporary Economics* (vol. II, pp. 183-222), Homewood, Illinois: Richard D. Irwin, Inc.
- Pigou, A.C. (1922). 'Empty Economic Boxes: A Reply', *The Economic Journal*, 32 no. 128 December: 458-65.
- Pigou, A. C. (1928). 'An Analysis of Supply', *The Economic Journal*, 38 no. 150 June: 238-57.

- Pollard, Sidney (1968). *The Genesis of Modern Management*, London: Penguin.
- Rees, R. (1974). 'A Reconsideration of the Expense Preference Theory of the Firm', *Economica*, 41 no. 163 August: 295-307.
- Ridley, Matt (2010). *The Rational Optimist: How Prosperity Evolves*, New York: HarperCollins Publishers.
- Rizvi, S. Abu Turab (2006). 'The Sonnenschein-Mantel-Debreu Results after Thirty Years', *History of Political Economy*, 38(Suppl 1): 228-45.
- Robbins, Lionel (1935). *An Essay on the Nature & Significance of Economic Science*, 2nd edn., Revised and Extended, London: Macmillan and Co. Limited.
- Roberts, John (2004). *The Modern Firm: Organizational Design for Performance and Growth*, Oxford: Oxford University Press.
- Robinson, E. A. G. (1931). *The Structure of Competitive Industry*, London: Nisbet & Co.
- Robinson, Joan (1933). *The Economics of Imperfect Competition*, London: Macmillan and Co., Ltd.
- Roe, Mark J. (1994). *Strong Managers, Weak Owners: The Political Roots of American Corporate Finance*, Princeton: Princeton University Press.
- Rothchild, (1947). 'Price Theory and Oligopoly', *The Economic Journal*, 57 no. 227 September: 299-320.
- Salanié, Bernard (2005). *The Economics of Contracts: A Primer*, 2nd edn., Cambridge, Mass.: The MIT Press.
- Sappington David E. and Joseph E. Stiglitz (1987). 'Privatization, Information and Incentives', *Journal of Policy Analysis & Management*, 6(4) Summer: 567-85.
- Sautet, Frederic E. (2000). *An entrepreneurial theory of the firm*, London and New York: Routledge.
- Sawyer, Malcom C. (1979). *Theories of the Firm*, London: Weidenfeld and Nicolson.
- Schmidt, Klaus (1996a). 'Incomplete Contracts and Privatization', *European Economic Review*, 40(3-5): 569-79.
- Schmidt, Klaus (1996b). 'The Costs and Benefits of Privatization: An Incomplete Contracts Approach', *The Journal of Law, Economics & Organization*, 12(1): 1-24.
- Shackle, G. L. S. (1967). *The Years of High Theory: Invention and Tradition in Economic Thought 1926-1939*, Cambridge: Cambridge University Press.
- Shafer, W. and H. Sonnenschein (1982). 'Market Demand and Excess Demand Functions'. In K. Arrow and M. Intriligator (eds.), *Handbook of Mathematical Economics* (vol. 2, pp. 671-93), Amsterdam: North-Holland.
- Shapiro, Carl and Robert D. Willig (1990). 'Economic Rationales for Privatization in Industrial and Developing Countries'. In Ezra N. Suleiman and John Waterbury, (eds.), *The Political Economy of Public-Sector Reform and Privatization*, Boulder: Westview Press.
- Shelanski, Howard A. and Peter G. Klein (1995). 'Empirical Research in Transaction Cost Economics: A Review and Assessment', *Journal of Law, Economics & Organization*, 11(2) October: 335-61.
- Shleifer, Andrei and Robert W. Vishny (1994). 'Politicians and Firms', *Quarterly Journal of Economics*, 109(4) November: 995-1025.

- Shoven, John B. and John Whalley (1984). 'Applied General-Equilibrium Models of Taxation and International Trade: An Introduction and Survey', *Journal of Economic Literature*, 22(3) September: 1007-51.
- Sidgwick, Henry (1901). *The Principles of Political Economy*, 3rd edn., London: Macmillan and Co. Limited.
- Silver, Morris (1995). *Economic Structures of Antiquity*, Westport Connecticut: Greenwood Press.
- Simon, Herbert A. (1951). 'A Formal Theory of the Employment Relationship', *Econometrica*, 9(3) July: 293-305.
- Simon, Herbert A. (1955). 'A Behavioral Model of Rational Choice', *The Quarterly Journal of Economics*, 69(1) February: 99-118.
- Simon, Herbert A. (1962). 'New Developments in the Theory of the Firm', *American Economic Review*, 52(2) May: 1-15.
- Smith, Adam (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*, Volumes I and II, R. H. Campbell and A. S. Skinner (general eds.), W. B. Todd (textual ed.), Indianapolis: Liberty Classics, 1981.
- Sobel, Robert (1999). *The Pursuit of Wealth: The Incredible Story of Money Throughout the Ages*, New York: McGraw-Hill.
- Solow, Robert (1971). 'Some Implications of Alternative Criteria for the Firm'. In R. Marris and A. Wood, (eds.), *The Corporate Economy: Growth, Competition, and Innovative Potential* (pp. 318-42), London: Macmillan.
- Spulber, Daniel F. (2008). 'Discovering the Role of the Firm: The Separation Criterion and Corporate Law', Northwestern Law & Economics Research Paper No. 08-23, December 6.
- Spulber, Daniel F. (2009). *The Theory of the Firm: Microeconomics with Endogenous Entrepreneurs, Firms, Markets, and Organizations*, Cambridge: Cambridge University Press.
- Sraffa, Piero (1926). 'The Laws of Returns under Competitive Conditions', *The Economic Journal*, 36 no. 144. December: 535-50.
- Stigler, George J. (1951). 'The Division of Labor is Limited by the Extent of the Market', *Journal of Political Economy*, 59(3) June: 185-93.
- Sue Wing, Ian (2004). 'Computable General Equilibrium Models and Their Use in Economy-Wide Policy Analysis: Everything You Ever Wanted to Know (But Were Afraid to Ask)', MIT Joint Program on the Science and Policy of Global Change, Technical Note No. 6, September.
- Sweezy, Paul M. (1939). 'Demand Under Conditions of Oligopoly', *Journal of Political Economy*, 47(4) August: 568-73.
- Tadelis, Steven and Oliver E. Williamson (2012). 'Transaction Cost Economics', March 12. Available at SSRN: <http://ssrn.com/abstract=2020176>
- Tirole, Jean (1988). *The Theory of Industrial Organization*, Cambridge Mass.: MIT Press.
- Tirole, Jean (2006). *The Theory of Corporate Finance*, Princeton: Princeton University Press.
- Tudge, Colin (1998). *Neanderthals, Bandits and Farmers: How Agriculture Really Began*, New Haven: Yale University Press.
- Vannoni, Davide (2002). 'Empirical Studies of Vertical Integration: The Transaction Cost Orthodoxy', *RISEC: International Review of Economics and Business*, 49(1) March: 113-41.

- Vickers, John and George Yarrow (1988). *Privatization: An Economic Analysis*, Cambridge, Mass.: MIT Press.
- Viner, Jacob (1937). *Studies in the Theory of International Trade*, New York: Harper and Brothers Publishers.
- Walker, Paul (2010). 'The (non)theory of the knowledge firm', *Scottish Journal of Political Economy*, 57(1) February: 1-32.
- Walker, Paul (forthcoming). 'The 'Reference Point' Approach to the Theory of the Firm: An Introduction', *Journal of Economic Surveys*.
- Walsh, Aidan (2009). 'A Mengerian theory of the origins of the business firm', *Studies in Emergent Order*, 2: 54-75.
- Wernerfelt, Birger (1984). 'A Resource-based View of the Firm', *Strategic Management Journal*, 5(2) April-June: 171-80.
- Wernerfelt, Birger (1997). 'On the Nature and Scope of the Firm: An Adjustment Cost Theory', *Journal of Business*, 70(4) October: 489-514.
- Whinston, Michael D. (2003). 'On the Transaction Cost Determinants of Vertical Integration', *Journal of Law, Economics, & Organization*, 19(1)April: 1-23.
- Wicksteed, Philip H. (1894). *An Essay on the Co-ordination of the Laws of Distribution*, 1932 edition, Reprint No. 12, London: London School of Economics.
- Williams, Philip L. (1978). *The Emergence of the Theory of the Firm: From Adam Smith to Alfred Marshall*, London: The Macmillan Press.
- Williamson, Oliver E. (1964). *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm.*, Englewoods Cliffs, N.J.: Prentice-Hall..
- Williamson, Oliver E. (1970). *Corporate Control and Business Behavior*, Englewoods Cliffs, N.J.: Prentice-Hall..
- Williamson, Oliver E. (1971). 'The vertical integration of production: market failure considerations', *American Economic Review*, 61(2) May: 112-23.
- Williamson, Oliver E. (1973). 'Markets and hierarchies: some elementary considerations', *American Economic Review*, 63(2) May: 316-25.
- Williamson, Oliver E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: The Free Press.
- Williamson, Oliver E. (1977). 'Firms and Markets'. In Sidney Weintraub (ed.), *Modern Economic Thought* (pp. 185-202), Philadelphia: University of Pennsylvania Press.
- Williamson, Oliver E. (1979). 'Transaction-Cost Economics: The Governance of Contractual Relations', *Journal of Law and Economics*, 22(2) October: 233-61.
- Williamson, Oliver E. (1985). *The Economic Institutions of Capitalism*, New York: The Free Press.
- Williamson, Oliver E. (1993). 'Introduction'. In Oliver E. Williamson and Sidney G. Winter (eds.), *The Nature of the Firm: Origins, Evolution, and Development* (pp. 3-17), New York, Oxford: Oxford University Press.
- Williamson, Oliver E. (1996a). *The Mechanisms of Governance*, Oxford: Oxford University Press.
- Williamson, Oliver E. (1996b). 'Transaction cost economics and the Carnegie connection', *Journal of Economic Behavior & Organization*, 31(2) November: 149-55.

- Williamson, Oliver E. (2000). 'The New Institutional Economics: Taking Stock, Looking Ahead', *Journal of Economic Literature*, 38(3) September: 595-613.
- Witt, Ulrich (1999). 'Do Entrepreneurs Need Firms? A Contribution to a Missing Chapter in Austrian Economics', *Review of Austrian Economics*, 11(1-2): 99-109.
- Woodland, Alan (2011). 'General Equilibrium Trade Theory'. In Daniel Bernhofen, Rod Falvey, David Greenaway and Udo Kreickemeier (eds.), *Palgrave Handbook of International Trade* (pp. 39-87), Houndmills, Basingstoke, Hampshire, U.K.: Palgrave Macmillan.
- Yu, Tony Fu-Lai (1999). 'Toward a Praxeological Theory of the Firm', *Review of Austrian Economics*, 12(1): 25-41.
- Zame, William R. (2007). 'Incentives, Contracts, and Markets: A General Equilibrium Theory of Firms', *Econometrica*, 75(5) September: 1453-500.
- Zenger, Todd R., Teppo Felin and Lyda Bigelow (2011). 'Theories of the Firm—Market Boundary', *The Academy of Management Annals*, 5(1) June: 89-133
- Zingales, Luigi (2000). 'In Search of New Foundations', *Journal of Finance*, 55(4) August: 1623-53.