

# THE (NON)THEORY OF THE KNOWLEDGE FIRM

*Paul Walker\**

## ABSTRACT

*This paper argues that the mainstream approaches to the theory of the firm do not provide a theory of the human capital-based or knowledge firm. We examine the neoclassical theory of the firm, the transaction cost model, the incentive-system approach and the Grossman Hart Moore approach and argue that none of them is able to fully capture the changes to the firm that the movement towards a knowledge economy entails. We also consider the effects of knowledge on the organisation of production. Will production take place within a single large factory, or several smaller factories or even within households?*

## I INTRODUCTION

John Kay makes the point that '[i]t is a cliché that we live today in a knowledge economy' (Kay, 2005, p. 266).<sup>1</sup> In recent years, it has become common for politicians and commentators to argue that changes in technology, in particular information and communication technology (ICT), have become the major driver of economic growth, so much so that many commentators argue that we are now in a 'new economy'.<sup>2</sup> For Alcaly (2003), the new economy developed in response to pressures from the application of information technologies in conjunction with increased global competition, deregulation and financial innovation. These factors have altered the whole business environment. Change is affecting not just markets but also firms within those markets.<sup>3</sup> Rajan and

\*University of Canterbury

<sup>1</sup> Sometimes also called the information economy, the digital economy, or the weightless economy. For more on the new economy see, for example, Alcaly (2003), Brynjolfsson and Kahin (2000), Coyle (2001), Oxley *et al.* (2008) and Shapiro and Varian (1999).

<sup>2</sup> Alcaly (2003, p. 4), for example, argues that:

... much is new about this new economy, particularly its signature information technology, the broad combination of technical equipment and know-now that enables us to process, store, and transmit information more efficiently. There have also been significant changes in the ways businesses operate, in the extent of trade and economic integration among nation – globalization – an in the influence and inventiveness of financial markets, including the stock and junk-bond markets.

<sup>3</sup> Spulber (2008, p. 5, footnote 8) gives the origin of the word 'firm' as:

The word 'firm' derives from the Latin word 'firmare' referring to a signature that confirmed an agreement by designating the name of the business.

Zingales (2003, p. 80) argue that the combination of increased competition, changes in technology and widespread access to finance have reduced the advantages of the large, vertically integrated firm. Thus, they say, we should expect to see the largest firms shrinking and this, they claim, is in fact what we have seen. They go on to point out that these same factors have also affected the relationship between physical and human capital in firm's creation of value and this is changing the organisational structure of the firm.

Human capital is replacing inanimate assets as the most important source of corporate capabilities and value. In both their organizational structure and their promotion and compensation policies, large firms are becoming more like professional partnerships.

(Rajan and Zingales, 2003, p. 90)

Further, Rajan and Zingales (2003, p. 87) argue that we are in fact seeing a new 'kinder, gentler firm'. This is in response to the changing balance of power within firms driven by the increasing importance of the worker.

That firms are changing matters because firms are the institutional structure within which most economic activity takes place and so, as they change much of our economic lives change. Simon (1991, p. 27) makes this point about the importance of firms by asking us to envision a mythical visitor from Mars. This visitor approaches Earth from space equipped with a telescope that reveals social structures. What our visitor's telescope would display is, according to Simon, that firms are the 'dominant feature of the landscape'. For Simon, the view that our visitor would get of the developed world or parts of the old Soviet Union or areas of urban China or urban India is one where most of the economic activity takes place, not within markets or self-sufficient households, but within the framework of firms.<sup>4</sup> But despite the obvious empirical importance of the firm, when our visitor turns its attention to an overview of the theory that purports to explain this 'dominant feature of the landscape', its survey would reveal a somewhat peculiar looking theoretical terrain. As Hart (1989, p. 1757) has written<sup>5</sup>:

<sup>4</sup> Roberts (2004, pp. 77–8) explains that:

In fact, John McMillan (2002, pp. 168–9) estimates that less than a third of all the transactions in the U.S. economy occur through markets, and instead over 70 percent are within firms.

<sup>5</sup> In recent personal correspondence (November 2008 – used here with permission), Professor Hart said of the 1989 quote:

The 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.

A catalogue of the progress referred to would include Alonso *et al.* (2008), Bolton and Dewatripont (1994), Cremer *et al.* (2007), de Meza and Lockwood (1998), Dessein and Santos (2003), Garicano (2000), Hart (2007), Hart and Holmström (2008), Hart and Moore (2007), Holmström (1999), Holmström and Milgrom (1991), Holmström and Milgrom (1994), Rajan and Zingales (1998) and Wernerfelt (1997).

An 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.

While Hart's point is made with reference to the orthodox view of the theory of the physical capital-based firm, it applies with even greater vengeance when we consider the human capital-based firm relevant to the knowledge or information economy. None of the textbook (neoclassical) model, the transaction cost model, the incentive-system approach and the Grossman Hart Moore (GHM) approach to the firm is able to fully capture the changes to the firm that the movement towards a knowledge economy entails. As knowledge becomes more important in the economy, human capital becomes more important to the firm and physical capital relatively less so. The major asset of a knowledge firm is its workers' human capital. Crucially, this increases the workers' importance and thus improves their outside options and changes the power relationships within the firm. Firms' organisational structures are changing to reflect this new reality. On the other hand, the orthodox theories of the firm are, in the main, silent about the changes that this increase in the importance of human capital is bringing about. This essay will examine these theories in an attempt to delineate the reasons for this silence.

The most deafening silence comes from the neoclassical model. And for good reason. The model of the 'firm' found in most microeconomic textbooks is not a 'theory of the firm' in any meaningful sense. The output side of the neoclassical model is a theory of supply rather than a genuine theory of the firm. Many would date the beginning of a serious theory of the firm as recently as Knight (1921) or Coase (1937). Before 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'. As has been pointed out by Demsetz (1982, 1988, 1995), before Knight and Coase, the fundamental preoccupation of economists was with the price system and hence little attention was paid to the firm as an important economic entity.

The interest in the price system, culminating in the 'perfect competition' model, has its intellectual origins in the 18th-century debate between free traders and mercantilists. This debate was, to a large degree, about the proper scope of government in the economy and the model it gave rise to reflect 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'. 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, pp. 27–8). For economists, the 200 years following Smith involved a search for conditions under which the price system would not descend into chaos.

The formal model that arose from this search is one that abstracts from any form of centralised control in the economy.<sup>6</sup> It is a model delineated by ‘perfect decentralisation’. Authority, be it in the form of a government or a firm or a household, plays no role in coordinating resources. The only parameters guiding decision making are those given within the model – tastes and technologies – and those determined impersonally on markets – prices. All parameters are outside the control of any of the economic agents and this effectively deprives all forms of authority a role in resource allocation.

In neoclassical theory, the firm is a ‘black box’ there to explain how changes in inputs lead to changes in outputs. 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, p. 9), the definition of a firm in the neoclassical model:

What 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.

As inputs are combined in the optimal fashion by the actions of independent owners of inputs motivated solely by market prices, there is no need for ‘management of some by others’, there is 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 and as there are no physical assets controlled by the ‘firm’, there are no (residual) control rights over these assets to allocate. This means that there is no role for owners of a firm.

<sup>6</sup>For Smith, this would be an abstraction too far. Smith knew the importance of institutions to the proper functioning of the market economy. For a discussion of Smith’s thought see Kennedy (2005, 2008).

Within the neoclassical model of the price system, questions about the existence, definition and boundaries of the firm are to a large degree meaningless since firms, by any meaningful definition of that term, do not really exist. As Foss *et al.* (2000, p. 632) summarise it:

The 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.

Given there is no serious modelling of the firm, there is no way to deal with the knowledge firm within this framework. There are no organisational problems or any internal decision-making process, in fact, there is no organisational structure at all and thus the advent of the knowledge economy cannot alter this nonexistent structure. As there is no role for managers or employees, there can be no role for human capital in the firm. But as was noted above, human capital is increasingly becoming the major creator of value in real firms.

The rest of this essay proceeds as follows. The contemporary mainstream approaches to the firm, and their shortcomings with regard to the human capital-based firm, are analysed in Sections II–IV. Section II covers the transaction cost approach, Section III looks at the incentive-system theory while Section IV considers the GHM model along with two extensions. Section V looks at the effects that knowledge and information have on the location of production, Section VI briefly covers related parts of the management literature and Section VII is the conclusion.

## II THE TRANSACTION COST APPROACH

The recognition that writing a contract is costly lies at the heart of the large and growing literature on the transaction cost approach to the firm. This literature has been developed by, among others, Williamson (1975, 1985, 1996) and Klein *et al.* (1978). Coase (1937) started the transaction cost approach by making the simple but important point: that there are costs to carrying out market transactions. Costs, which today are called transaction costs. Coase describes such costs as:

What the prices are have to be discovered. There are negotiations to be undertaken, contracts have to be drawn up, inspections have to be made, arrangements have to be made to settle disputes, and so on.

(Coase, 1992, p. 715)

Roberts (2004, p. 90) defines transaction costs as:

... the costs of finding and qualifying trading partners, of establishing specifications and prices, of negotiating and drafting contracts, and of monitoring and enforcing agreements. They are also the opportunity costs of lost benefits that are occasioned by the difficulties of developing complete, enforceable agreements between separate parties.

Transaction cost economics is based on the ideas of bounded rationality – intendedly rational, but only limitedly so – and opportunism – self-interest with guile. Two consequences of these assumptions are that contracts will be incomplete and contracts may not be honoured. Hart (1995, p. 23) argues that in transaction cost economics, contracts are incomplete for three reasons, all of which are, in his view, forms of bounded rationality:

First, in a complex and highly unpredictable world, it is hard for people to think very far ahead and to plan for all the various contingencies that may arise. Second, even if individual plans can be made, it is hard for the contracting parties to negotiate about these plans, not least because they have to find a common language to describe states of the world and actions with respect to which prior experience may not provide much of a guide. Third, even if the parties can plan and negotiate about the future, it may be very difficult for them to write their plans down in such a way that, in the event of a dispute, an outside authority – a court, say – can figure out what these plans mean and enforce them. In other words, the parties must be able to communicate not only with each other, but also with outsiders who may have little knowledge about the environment in which the contracting parties operate.

But why do incomplete contracts matter? If parties to a contract can renegotiate the contract, and thus fill in any gaps, why is contractual incompleteness an issue? First, there may be costs to haggling over the terms and conditions of the new contract. Haggling over the division of surpluses is inefficient in that it is time-consuming and wastes resources while serving no productive purpose. Second, informational asymmetries may prevent the parties from reaching an efficient outcome. Assume the buyer of an input does not know the actual cost of the input but only knows the probability distribution from which the costs are drawn. The seller of the input knows the true cost. Supply of the input can be ensured by a high price offer from the buyer. If the buyer wants to cover the seller's costs with probability one, then this could be an expensive option as the buyer will be overpaying in the low-cost states of the world. If a low price offer is made, then the seller will not supply in the high-cost states of the world and so profit maximising behaviour by the buyer may lead to profitable trades not taking

place.<sup>7</sup> An important point here is that if switching to a new trading partner at the renegotiation stage was easy, then neither of these two costs would be significant. Thus for these costs to be high, there must be something preventing the switching to a new trading partner. That ‘something’ is normally taken to be *ex ante* relationship-specific investment. In other words, a prior investment whose value is greatest when the contracting parties relationship extends over time but for which little or no value is created if the parties relationship breaks down. Such investments are normally thought of as investments in physical capital but as Roberts (2004, p. 91) points out:

Firm-specific human capital – knowledge that is only (or especially) valuable in the context of employment with a particular firm – is another example.

Relationship-specific investments result in there being a third cost of incomplete contracts. It may be that because contracts are incomplete, parties are deterred from making efficient relationship-specific investments. In a comprehensive contracts world, relationship-specific investments could be protected by enforceable contracts. In an incomplete contracts world, this may not be possible. Parties will recognise that any long-term contract between them will be incomplete. This could be because of problems such as being unable to specify far in advance the quality and quantity of the goods traded. This incompleteness will mean the contract will be subject to renegotiation. Even in a situation where problems of haggling and asymmetric information do not arise, the gains from trade will have to be divided and this division will depend on the *ex post* bargaining strengths of the parties rather than on what is written in the contract or on the grounds of economic efficiency. This raises the fear that one party could be exploited by another. For example, an input supplier, who has made (sunk) relationship-specific investments, may worry that the buyer will take advantage of these investments to drive the price he pays for the input down to around variable cost, so there is little or no contribution to covering the investment costs. But it is still not worthwhile for the supplier to stop supplying. This is simply because the sunk investment costs have to be paid whether or not supply takes place, and the asset has no other profitable use. This exploitation of a quasi-rent – returns greater than what is required to keep an asset in its current use once it has been created – is the classic ‘hold-up’ problem.

Realising that such exploitation could occur may result in the supplier being unwilling to undertake the investment in the first place. Thus the buyer, if he wants the input supplied at all, may have to produce it himself. The buyer could purchase the supplier, i.e. vertically integrate with the supplier, thereby making the supplier part of the buyer’s firm. This eliminates the hold-up problem since the quasi-rents now accrue to the buyer. Use of the investment asset is now directly under the control of the buyer and all costs of and benefits from

<sup>7</sup> Assume the buyer values the input at 1. The seller’s costs are  $1/2$  or 1, each with probability  $1/2$ . An offer by the buyer of 1 ensures supply in all states of the world but results in zero profits for the buyer. An offer of  $1/2$  results in non-supply in the high-cost states of the world but profits of  $1 - 1/2 = 1/2$  in the low-cost states of the world. Thus, the low price (trade only half the time) offer is more profitable than the high price (trade at all times) offer and therefore profitable trades do not take place.

investment have been internalised. The investment decision is now just part of the buyers profit maximisation problem.

This argument that vertical integration deals with the hold-up problem is strongest when the assets involved are physical. The argument is less applicable to relationship-specific investments in human capital. As human capital cannot be owned by anyone other than the particular individual acquiring it, the potential for opportunistic behaviour still exists even after vertical integration. The buyer does not have control over the human capital in the way he does over the physical capital. The individual who invests in the relationship-specific human capital still controls that capital even after becoming part of the buyer's firm and thus they can still hold-up the buyer. Thus, the explanation for the existence of firms as the answer to hold-up problems, related to relationship-specific investments, does not apply in the case of a human capital only firm.

This point has been commented on by Klein (1988, p. 204) in his discussion of the (in)famous case of the vertical integration of General Motors with Fisher Body.<sup>8</sup> It is noted that vertical integration did not eliminate the Fisher brothers, they just went from being an independent contractor to being employees of General Motors. Given that human capital cannot be owned, Klein asks '... how did the vertical integration of General Motors with Fisher reduce the hold-up problem?' (Klein, 1988, p. 206). He argues that even post-integration, the Fisher brothers could still hold up General Motors, in this case for their human capital-specific investments.

### III THE INCENTIVE-SYSTEM THEORY

This approach to the theory of the firm was developed by Holmström and Milgrom (1991, 1994), Holmström and Tirole (1991) and Holmström (1999) and has been described by Gibbons (2005, p. 206) as an 'accidental theory of the firm'. The reason for Gibbons's description is that the focus of these papers was not on the make-or-buy problem of the transaction cost or GHM approaches but rather on a multi-task, multi-instrument principal-agent problem and its application to the firm was an 'accidental' outcome of this endeavour.

To analyse the application of this theory to the knowledge firm, we will take advantage of Gibbons (2005, pp. 210–2) 'stick-figure rendition' of the theory. In the simple Gibbons model, there is a technology of production, which is a linear combination of the agent's actions:  $y = f_1 a_1 + f_2 a_2 + \varepsilon$ , where  $a_1$  and  $a_2$  are actions chosen by the agent and  $\varepsilon$  is a noise term. Evaluation of performance by the agent is based on an indicator  $p$ , which is a different linear combination of the agent's actions:  $p = g_1 a_1 + g_2 a_2 + \phi$ , where  $\phi$  is another noise term. Gibbons assumes that both parties are risk-neutral,  $\omega$  is the total compensation paid by the principal to the agent and  $c(a_1, a_2)$  represents the agent's cost function. Gibbons makes the assumption that

$$c(a_1, a_2) = \frac{1}{2}a_1^2 + \frac{1}{2}a_2^2.$$

<sup>8</sup>For more on the General Motors-Fisher Body integration, see Baird (2003), Casadesus-Masanell and Spulber (2000), Coase (2000, 2006), Freeland (2000), Goldberg (2008), Klein (1988, 1996, 2000, 2007, 2008) and Klein *et al.* (1978).

In addition, Gibbons assumes that the principal and the agent sign a linear contract,  $\omega = s + bp$ , based on the performance indicator  $p$ .

To provide a theory of the firm, this model has to be extended to include physical capital, a machine, which is used by the agent during the production of  $y$ . After production this capital has a value determined by a third linear combination of the agent's actions:  $v = h_1 a_1 + h_2 a_2 + \xi$ , where  $\xi$  is a third noise term. The choice variables in the model are therefore the agent's actions  $a_i, i = 1, 2$  and  $b$  the slope of the optimal contract. As a point of comparison, note that the first-best actions of the agent are those that maximise the expected total surplus, that is, they will maximise the expected value of the sum of the principal's payoff,  $y - \omega$ , the agent's payoff,  $\omega - c(a_1, a_2)$ , and the value of the physical asset,  $v$ .

$$\begin{aligned} TS^{FB} &= E(y - \omega + \omega - c(a_1, a_2) + v) \\ &= E(y + v) - c(a_1, a_2) \\ &= E(f_1 a_1 + f_2 a_2 + \varepsilon + h_1 a_1 + h_2 a_2 + \xi) - c(a_1, a_2) \\ &= f_1 a_1 + f_2 a_2 + h_1 a_1 + h_2 a_2 - c(a_1, a_2) \text{ assuming } E(\varepsilon) = E(\xi) = 0 \\ &= f_1 a_1 + f_2 a_2 + h_1 a_1 + h_2 a_2 - \left( \frac{1}{2} a_1^2 + \frac{1}{2} a_2^2 \right) \end{aligned}$$

and therefore  $a_1^{FB} = f_1 + h_1$  and  $a_2^{FB} = f_2 + h_2$ .<sup>9</sup>  $TS^{FB}$  is independent of the value of  $b$ .

If the principal owns the machine, then the agent is an employee of his firm and the principal's payoff is  $y + v - \omega$ , while the agent's payoff is  $\omega - c$ . In this case, the agent's optimal actions maximise

$$\begin{aligned} E(\omega) - c(a_1, a_2) &= E(s + bp) - \left( \frac{1}{2} a_1^2 + \frac{1}{2} a_2^2 \right) \\ &= E(s + b(g_1 a_1 + g_2 a_2 + \phi)) - \left( \frac{1}{2} a_1^2 + \frac{1}{2} a_2^2 \right) \\ &= s + b g_1 a_1 + b g_2 a_2 - \left( \frac{1}{2} a_1^2 + \frac{1}{2} a_2^2 \right) \text{ assuming } E(\phi) = 0. \end{aligned}$$

The optimal actions are therefore,  $a_{1E}^*(b) = b g_1$  and  $a_{2E}^*(b) = b g_2$ .<sup>10</sup> The efficient contract slope,  $b_E^*$ , maximises the expected total surplus,  $E(y + v) - c(a_1, a_2)$  or  $TS_E(b) = (f_1 + h_1) a_{1E}^*(b) + (f_2 + h_2) a_{2E}^*(b) - \left( \frac{1}{2} a_{1E}^*(b)^2 + \frac{1}{2} a_{2E}^*(b)^2 \right)$ .

$$\begin{aligned} &^9 \frac{\partial TS^{FB}}{\partial a_i} = 0 \\ &\Rightarrow f_i + h_i - a_i = 0 \\ &\Rightarrow a_i^{FB} = f_i + h_i \end{aligned}$$

$$\begin{aligned} &^{10} \frac{\partial E(\omega) - c(a_1, a_2)}{\partial a_i} = 0 \\ &\Rightarrow b g_i - a_i = 0 \\ &\Rightarrow a_{iE}^*(b) = b g_i \end{aligned}$$

Alternatively, the machine can be owned by the agent. Gibbons interprets this case as the agent being an independent contractor. In this situation, the payoffs for the principal will be  $y - \omega$  and for the agent they are  $\omega + v - c$ . The optimal actions for the agent will therefore be,  $a_{1C}^*(b) = g_1 b + h_1$  and  $a_{2C}^*(b) = g_2 b + h_2$ .<sup>11</sup> For this case, the efficient slope,  $b_C^*$ , will maximise the expected total surplus of

$$TS_C(b) = (f_1 + h_1)a_{1C}^*(b) + (f_2 + h_2)a_{2C}^*(b) - \left( \frac{1}{2}a_{1C}^*(b)^2 + \frac{1}{2}a_{2C}^*(b)^2 \right).$$

Gibbons (2005, p. 211) summarises the analysis so far as:

... having the agent own the asset causes the agent to respond to a given contract slope ( $b$ ) differently than when the agent does not own the asset,<sup>12</sup> so the make-or-buy problem amounts to determining which of the agent's best-response functions – that of the employee,  $(a_{1E}^*(b), a_{2E}^*(b))$ , or that of the independent contractor,  $(a_{1C}^*(b), a_{2C}^*(b))$  – allows the parties to achieve greater total surplus.

The discussion so far has relied on an implicit assumption that the value of the asset is not contractible and therefore the owner of the asset receives its value. Since the asset's value is not contractible, putting ownership in the hands of the agent provides him with incentives that cannot be replicated via a contract. But providing the agent with the incentive to increase the value of the asset may or may not help the principal control the agent's incentives via contract. That is, if the agent owns the asset, he has two sources of incentives, the asset's post-production value and the contracted for performance. Without ownership, he concentrates solely on the contracted for performance. Integration would be efficient, that is, having the principal own the asset is efficient, when having the agent do so hurts the principal's efforts to create incentives via contract.

When we turn to consider the case where the additional capital in the model is not a machine but is human capital, as would be important in a knowledge firm, an important difference arises: ownership can no longer be transferred, as it can in the physical capital case. If the asset is a human capital asset, then, without slavery, ownership must remain with the person who has made the investment. Also, as above, the non-contractibility of the asset's value means the principal and the agent cannot transfer the value of the asset as part of a contract. So in contrast to the physical asset case, with a human asset, neither ownership nor value can be transferred. Thus, if the agent (principal) makes an investment, the

<sup>11</sup>  $E(\omega + v) - c(a_1, a_2) = E(s + b(g_1 a_1 + g_2 a_2 + \phi) + h_1 a_1 + h_2 a_2 + \xi) - \left( \frac{1}{2}a_1^2 + \frac{1}{2}a_2^2 \right)$   
 $= s + b g_1 a_1 + b g_2 a_2 + h_1 a_1 + h_2 a_2 - \left( \frac{1}{2}a_1^2 + \frac{1}{2}a_2^2 \right)$   
 assuming  $E(\phi) = E(\xi) = 0$ .

<sup>12</sup> That is,  $a_{1E}^*(b) \neq a_{1C}^*(b)$

value of the investment stays with the agent (principal) irrespective of the form of the relationship between the principal and the agent. Importantly, asset ownership can no longer determine whether the agent is an employee or a contractor. Therefore, a human asset cannot serve as an instrument in the incentive problem in the same way as a physical asset. This means that the incentive-system theory (as described above) cannot act as a model of a human capital-based firm.

#### IV THE GHM APPROACH

Another of the more recent theories of the firm is based on the works of Grossman and Hart (1986, 1987) and Hart and Moore (1990).<sup>13</sup> Within the GHM approach, ownership is defined in terms of residual control over non-human assets, things such as machinery, inventories, buildings, patents, client lists, firm's reputation, etc. Owner-managers employ labour who cannot work without the physical capital these firms own. Dismissal/resignation of the labour requires them to find other physical capital owning organisations (firms) to employ them. On liquidation of the firm, physical capital can be sold and the proceeds disbursed to the owners (shareholders). The standard GHM theory of the firm is based on the role of non-human capital in the firm. The definition of a firm, the determinants of the boundaries of a firm – that is, the determinants of vertical integration of firms, the meaning of ownership of the firm, the nature of authority within the firm are all functions of control rights over the firm's non-human assets. Making non-human assets the centre of the theory means that questions to do with the ownership and control of the physical information technology can be addressed but theory does not deal with firms based on human assets. However, it has been noted from the beginning that the theory could be extended to include human capital. As Hart (1988, p. 151) argues:

... one difference with previous work is the emphasis on how integration changes control over physical assets. This is in contrast to Coase's 1937 paper which focuses on the way integration changes an ordinary contractual relationship into one where an employee accepts the authority of an employer (within limits). Note that these approaches are not contradictory. Authority and residual rights of control are very close and there is no reason why our analysis of the costs and benefits of allocating residual rights of control could not be extended to cover human, as well as physical, assets.

But this extension is not entirely satisfactory.

Once we move to a situation where firms may own/need little physical capital, then the modern theory of the firm loses much of its purpose. Once human capital (labour) becomes the most important/sole creator of wealth/value added, modern economic theory is in need of modification. The theory does not,

<sup>13</sup> Introductions to this literature, of varying technical sophistication, can be found in Bolton and Dewatripont (2005, pp. 498–515), Hart (1988, 1989, 1995) and Moore (1992).

however, lose all relevance. As Hart (1995, pp. 56–7) explains, at least some, nonhuman assets are essential to a theory of the firm. To see why this may be so, consider a situation where ‘firm’ 1 acquires ‘firm’ 2, which consists entirely of human capital. The question Hart raises is: What stops firm 2’s workers from quitting? Without any physical assets – e.g. buildings – firm 2’s workers would not even have to relocate themselves physically. If these workers were linked by telephones or computers, which they themselves own, they could simply announce that they have decided to become a new firm. For the acquisition of firm 2 by firm 1 to make economic sense, there has to be a source of value in firm 2 over and above the human capital of the workers. It makes little sense to buy a ‘firm’ if that ‘firm’ can just get up and walk away. Hart argues there must be some ‘glue’ holding firm 2’s workers in place.

The value that acts as this glue may consist of as little as a place to meet; the firm’s name, reputation, or distribution network; the firm’s files containing important information about its operations or its customers; or even a contract that prohibits firm 2’s workers from working for competitors or from taking existing clients with them should they quit. The source of value may even just represent the difficulty firm 2’s workers face in coordinating a move to another firm. But, Hart points out, without something binding the firm together, the firm becomes a phantom, a flimsy and unstable entity, constantly subject to the possibility of break-up or dissolution.

Thus even a human capital-based firm will involve some nonhuman capital, but the human capital will play the dominant role. The important characteristic of human capital is that it embodies information and knowledge. A theory of the human capital-based firm has to model this co-existence of the human and nonhuman capital. Brynjolfsson (1994) deals with the issue by extending the property rights approach to the firm to include information whether this information is embodied in humans, in the form of human capital, or in artefacts. Rabin (1993) also works within the property rights framework but extends it by assuming that an agent has information, about how to make production more productive, which they are willing to sell. The problem is that if the information is not revealed before the agent is paid, a (potential) buyer may have little reason to believe the agent is truly well informed, but if the agent reveals the information up front, the buyer could simply use the information without payment. Rabin’s answer is that the informed agent gains control over productive assets and thus does not have to sell the information. We discuss Brynjolfsson (1994) and Rabin (1993) in more detail below.

The Brynjolfsson (1994) model considers an entrepreneur who has some expertise needed to run a firm. No value can be created without both the knowledge asset of the entrepreneur and the physical assets of the firm. He assumes that no comprehensive contract can be written between the entrepreneur and the firm. If the entrepreneur does not own the firm, then if he makes an investment in effort and creates value, he can be subject to hold-up by the other party since he needs the firm’s physical assets. If the entrepreneur owns the firm then clearly the hold-up problem ceases to exist. The most obvious

interpretation of Brynjolfsson model is as a model of a labour-owned firm. Brynjolfsson argues that it is optimal to give the entrepreneur ownership of the physical assets of the firm since he has information that is essential to its productivity. This result is obviously just an application of Hart and Moore's proposition that an agent who is 'indispensable' to an asset should own it.<sup>14</sup> Here, firms are owned by the indispensable human capital, or more normally by a small section of the human capital, e.g. a partnership. Labour-owned firms are one way to form a human capital-intensive firm but the shortcomings of such organisations are obvious: lack of access to capital, inadequate risk pooling, investment problems, older workers wanting a shorter pay-back period than younger workers are membership rights tradable and if so under what conditions, new members would have to purchase 'equity' in the business from retiring ones, borrowing to cover such a purchase could be a problem for younger would-be members, etc.

Rabin (1993) starts from the idea that there can be problems for an economic agent who wishes to sell potentially useful private information to others. Any person who believes that they have some valuable insights into how to produce something is unlikely to be able to exploit that information by offering to sell it to an existing firm. If they do not reveal their information before being paid, a firm may have little reason to believe the seller is truly well informed. If they reveal the information up front, the firm may simply use the information without compensating the would-be seller. When the current owner of the firm cannot observe how informed the seller is, Rabin argues that the information seller may have to buy the factory to make money from their private information. In fact, Rabin shows that the more bargaining power an informed party has, the more likely they are to obtain control of an asset. In Rabin's model, knowledge determines ownership of the firm; that is, the 'indispensable' human capital owns the physical capital just as in Brynjolfsson's model.

While Brynjolfsson's model is distinct from Rabin's model, they are complementary. The relationship between information, ownership and authority is central to both papers. Rabin works within a framework of an adverse selection model and shows that the adverse selection problems can be such that, in some cases, an informed party has to take over the firm to show that their information is indeed useful. The Brynjolfsson model is a moral hazard type framework that deals with the issue of incentives for an informed party to maximise non-contractible effort.

A potential problem is that part of what may be driving the results of both the Brynjolfsson and Rabin models is the implicit restriction of the GHM

<sup>14</sup> See Hart and Moore (1990, p. 1133). For Hart and Moore, an agent  $i$  is indispensable to an asset  $a_n$  if, without agent  $i$  in a coalition, asset  $a_n$  has no effect on the marginal product of investment for the members of that coalition. That is, for all agents  $j$  in any coalition  $S$  and for all sets  $A$  of assets containing  $a_n$ ,

$$v^j(S, A) \equiv v^j(S, A \setminus \{a_n\}) \quad \text{if } i \notin S.$$

framework that the owner is also the manager.<sup>15</sup> Hart and Holmström (2008, pp. 1–2) note that there are several ways to see this: (1) according to the GHM theory, the major impact of a change in ownership is on those who gain or lose ownership rights (the owners); however, in a merger between two large firms, it is often the case that the key decision makers (the managers) do not have substantial ownership rights before or after the merger. Hence, a model in which the decision makers lose or gain ownership rights must be a model where the owners are the managers; (2) the relationship-specific investments analysed are made by individuals rather than the firm; this applies better to small rather than large companies; and (3) the approach envisions a situation in which all the relevant parties meet *ex post* and bargain, to an efficient outcome using side payments, over the gains from trade and the only issue is who has the right to which assets; there are no other decisions in the model. As it stands, the model has no room for ‘organisational structure’, ‘hierarchy’ or ‘delegation’. In a sense, the model continues to describe a pure market economy, although one enriched by the idea that individuals can be empowered through the ownership of key nonhuman assets. Holmström (1999, p. 87) goes so far as to argue that the GHM framework is an owner–manager framework since the theory is a theory of asset ownership by individuals rather than by firms. ‘Assets are like bargaining chips in an entirely autocratic market’. There are neither firms nor workers! Ownership of assets by an individual only provides a theory of organisational identities if we associate the firms with those asset owning individuals.<sup>16</sup> The problem is that if human capital is important to the production process and it makes sense to allocate control of production to the informed party, that is, make that party the manager, then because the manager must also be the owner, we end up with the human capital owning the firm. Thus, a labour-owned firm may be an artefact of the model.

## V KNOWLEDGE AND PRODUCTION LOCATION

In his discussion of the neoclassical model Hart (1995, p. 17) notes that it tells us nothing about where a firm’s boundaries will lie or about the size or location of a plant or factory within a given firm. This approach is consistent both with every existing firm being a plant or division of one huge firm that produces everything and with every plant or division of each existing firm being a separate and independent firm in their own right. Thus, it is not clear in what organisational form production will occur. Will it be organised as a single large factory, several

<sup>15</sup> In addition, there appear to be technical issues with some of the results in the Rabin paper. In particular, it is not clear that the proof of Result 2.1 is valid which casts doubt on the result itself and the comparative static results that follow from it.

<sup>16</sup> Bolton and Dewatripont (2005, p. 515) also makes the point by noting:

Another limitation of the theory is that it is mainly a theory of entrepreneurial firms run by owner-managers. As we highlighted earlier, the theory predicts that ownership of assets by shareholders who are not in any way related to the underlying business of the firm is inefficient. Ownership should go only to agents who make *ex ante* investments, and it should not be shared.

smaller factories or a household? The GHM approach does delineate the boundaries of the firm but still does not tell us anything about the location or size of a plant or factory that is part of the firm. Again, the form of production organisation is indeterminate. What will be argued below is that knowledge and information are important influences on the form of organisation in which production takes place. The most obvious issue has to do with the determination of whether or not work occurs in centralised factories or in separate households or some combination of these. This has been an issue since at least the industrial revolution.

In his discussion of the development of the factory system during the industrial revolution, Mokyr (2001, 2002, Chapter 4) puts forward the argument that the location of production was dependant, in part, on the trade-off between 'the relative costs and benefits of moving people as opposed to moving information' Mokyr (2002, p. 120). That is, he develops a line of reasoning that suggests that one factor encouraging the organising of workers under a single roof, rather than in separate households, was the division of knowledge.<sup>17</sup> As long as there was little division of knowledge, so that the knowledge needed to carry out production could be summarised in a few basic rules, the household could know all that was needed to act as the 'unit of production'. The cost of transferring information between workers was low since there was little of it needed and the workers were contained within the household. Moving people between households or to factories, however, was slow and costly. But as technology developed, the competence required for production moved beyond the capability of a single household. As the knowledge needed to produce things increased, firms faced two related problems. First, a firm needed to be able to incorporate existing knowledge into their production system, and second, they had to be able to generate new knowledge to keep or establish a competitive advantage. Inevitably, specialisation and the division of labour became finer. The way to deal with the increased level of knowledge demanded for production was to divide up the production process into smaller manageable tasks. Workers had knowledge about ever smaller pieces of the production puzzle. A result of this more extensive division of labour, which could in some circumstances reinforce the movement towards a single location factory, was noted by Babbage (1835).<sup>18</sup> Babbage observed that the greater the division of labour the less time required for learning any requisite skills. This results in a lessening of the period during which a new entrant to the workforce would be relatively unproductive and unremunerative. Because less knowledge and training was required to learn

<sup>17</sup> Mokyr (2002, p. 131) outlines the three main explanations for the rise of the factory. 'One relies on fixed costs and technical and physical economies of scale and scope, which might have caused the minimum efficient size of plants to become larger than the household. A second explanation is drawn from the modern micro-economics of the firm: because of asymmetric information and the division of labor, costs were higher in decentralized households, and the new technology changed the benefits and costs of monitoring and the incentives to self-monitor. A third argument is that by concentrating all workers under one roof and placing them under supervision, actual labor effort is enhanced'. To these three Mokyr adds a fourth, the division of knowledge, which we discuss here.

<sup>18</sup> For more on Babbage and his work relevant to economics see Rosenberg (1994, Chapter 2).

to undertake a single operation, as opposed to that required to undertake many different operations, a new employee would more quickly reach a situation where he generates a profit for his employer.

Mokyr points out that the importance of the division of knowledge to the firm was first recognised, albeit in a non-historical setting, by Demsetz (1988) and formalised by Becker and Murphy (1992). What these works suggested was a new interpretation of the role of the firm. Given that there are limitations to what a worker can know, the competence that a firm has to possess to produce must be divided into manageable portions and allocated among the workers. The actions of the different groups of workers are then coordinated by the firm's management. Thus, workers who produce on the basis of knowledge they themselves do not possess have their activities directed by someone who does possess (at least more of) the necessary knowledge. Therefore, the coordination needed due to asymmetries in information among workers provides a rationale for management. In this way, direction is a substitute for education, that is, a substitute for the transfer of the knowledge itself. Specialisation in knowledge can, therefore, both exacerbate existing information asymmetries and create new ones. Any information asymmetry gives rise to an organisation problem for the firm: How can agents who possess knowledge be encouraged to reveal their knowledge fully and truthfully to other workers or management? Mokyr argues that:

Putting all workers under one roof ensured repeated interaction and personal contact provides maximal bandwidth to maximize the chances that the information will be transmitted fully and reliably. Inside a plant agents knew and could trust each other, and this familiarity turned out to be an efficient way of sharing knowledge.

(Mokyr, 2002, p. 141)

From this, it can be seen that as long as the minimum competence needed by a firm is small, the plant size can also be small and can, therefore, coincide with the household. When the competence needed grows, the unit of production has to change or an efficient network for knowledge distribution has to develop. At a time when the main technique for the distribution of knowledge was direct contact, as at the time of the industrial revolution, such networks in the form of professional associations of mechanics, machinists, engineers, etc., did develop. But the firm was also an answer to the problem of knowledge distribution. Costs of accessing knowledge were minimised in a single plant where workers could communicate face to face. Factories acted as repositories of technical knowledge and allowed workers to access this information at relatively low cost. Thus, factories allowed knowledge to pass in two directions: across space so that other workers could carry out a given task and through time so that knowledge passed from one generation of workers to the next.

McDermott (2001, p. 48) explains that the transition to factory life altered the incentives for both owners and workers with regard to acquiring knowledge and providing training. Workers now had the opportunity to acquire highly specialised know-how about both their particular firm's production process and about more general, transferable, skills. Owners now had an incentive to

educate their workforce since firms unlike households can go out of business if they do not keep up. This gave owners the needed impetus to train their employees. Galor and Moav (2006) also argue that support from capitalists for the education of workers was due to the increasing importance of human capital in sustaining their profits. Physical capital accumulation in the process of industrialisation gradually intensified the importance of skilled labour in the production process and thereby generated an incentive for investment in human capital. Because of the complementarity between physical and human capital in production, the capitalists were among the prime beneficiaries of the accumulation of human capital by workers. By putting workers under one roof, it was easier for owners to compare workers' productivity and select those who were most suitable for greater training and advancement.<sup>19</sup> This process helped differentiate the firm from the household.

Differences in knowledge between principals and agents can affect the desire to move to a centralised factory for additional reasons. Lamoreaux *et al.* (2003) note that the putting-out system in US cotton-spinning came under pressure, in part, because of principal-agent problems between the manufacturers and home-based weavers. Lamoreaux *et al.* (2003, pp. 412–3) write:

However, the enormous coordination problems that this system entailed (for example, unsupervised weavers working in their homes turning out fabrics of vastly varying qualities) spurred manufacturers to reconcentrate production in factories as soon as technological innovation in the form of the power loom enabled them to expand capacity sufficiently.

Kim (2001) suggests another way in which information can affect firm organisation and business location. Kim makes the point that specialisation adds to transaction costs via a loss in information. With specialisation, firms know their costs but are uncertain as to the demand for their products while consumers know their demands but not the supply conditions of production. Kim notes that before the late 19th century, most goods were produced using craft technology and were often produced in homes or at best small shops. Such firms tended to operate in a local or regional market, produced a single line of output, and were owned and managed by a single individual or a partnership. Given the relative simplicity of production, consumers could identify the quality level of products either via physical inspection or through the reputation of the producer or seller. As the production process became more sophisticated and production took place away from the consumers' location, information became more and more asymmetric. Producers knew the quality of their goods but

<sup>19</sup> Lamoreaux *et al.* (2003, p. 419) illustrate the basic idea with the example of the Ford Motor Company:

In mass-production enterprises where machines controlled the pace of work – the Ford Motor Company's assembly lines, for example – it was much easier to identify workers who were slacking off or unable to keep up. Hence Ford was able to secure a high level of effort from its work force by paying generous wages and dismissing workers who could not make the grade.

consumers were less well informed. This gave rise to a potential adverse selection, or 'lemons' problem, where bad products drive out good ones.

Kim's argument is that the modern multiunit firm is a solution to the asymmetric information problem.<sup>20</sup> Hence, production moved away from single-plant, single-region, single-product firms towards multi-plant, multi-region, multi-product enterprises to counter potential lemons problems. The advantage that the multiunit firm had was that as repeat sales were of greater value to the multiunit firm, they were better able to make large firm-specific, sunk cost investments in advertising and branding to credibly signal to buyers that the costs of renegeing on quality were high. Kim (2001, p. 311) writes:

Multiunit firms were able to solve the asymmetric information problem through the use of advertising and the development of brand names. In the presence of uncertain quality and the absence of a credible third party enforcer, the main private-contract enforcement mechanism relies on the value of repeat sales to a firm. One solution to signaling a firm's value of repeated sales is to invest in firm-specific and non-salvageable assets such as advertising and developing brand names. Since the value of repeat sales is limited for most single-unit firms, these firms have little incentive to advertise and develop brand names. On the other hand, for multiunit firms, the value of repeat purchase is potentially much greater. Thus, the economies of marketing for multiunit firms come not only from their ability to spread their costs over many plants or stores, but also from the fact that the cost of renegeing on their product quality is significantly higher.

Audia *et al.* (2001, pp. 79–82) see information affecting the location decision in other ways. For them, there is a trade-off in the organisation of production between geographic dispersion and organisational learning.<sup>21</sup> First, a

<sup>20</sup>The multiunit firm is defined as a firm that controls and manages from a central administrative organization the production decisions of establishments or plants in at least two different localities. Multiunit firms are usefully categorized into the following three types: horizontal, vertical, and conglomerate or diversified. Horizontal firms produce the same product in different locations, vertical firms use outputs of some of their plants as intermediate inputs to some of their other plants, and conglomerate or diversified firms manage plants in unrelated industries' (Kim, 1999, footnote 1).

<sup>21</sup>Audia *et al.* (2001, pp. 77–83) deal with other advantages and disadvantages of multiunit firms. Multiunit firms benefit from product-specific economies:

Multiunit firms achieve these efficiencies through product specialization. According to Scherer *et al.* (1975, p. 295) 'Product specialization exists when plants belonging to the same organization produce for a broad geographic market some narrow segment of the product line normally encompassed within an industry's definition.' Product specialization allows longer production runs that facilitate worker productivity, increase product quality and simplify production planning' (Audia *et al.* 2001, p. 77).

When a firm is made up of multiple units, the geographic distribution of its facilities can affect the firm's performance. Multi-local firms – firms operating in dispersed locations – can gain competitive advantages from several sources: (1) they can minimise transportation costs by producing close to customers. (2) Operating at different locations means a firm can weather idiosyncratic risks associated with a given area. (3) Operation in multiple markets could allow firms to collude with other multi-market competitors. A disadvantage of multiunit firms is bureaucratisation. Large administrative structures reduce the firms' ability to react to environmental changes.

multi-plant firm has advantages in the creation of new knowledge with regard to the optimal method of production. With multiple sites, even random variation across those sites can offer the opportunity to gather comparative information on the best production methods. But more strategically, multi-unit firms can undertake parallel experimentation with different plants undertaking different experiments at the same time. This offers at least two advantages: (1) it allows learning to occur at a faster pace. As experimentation takes place in chronological time, being able to undertake multiple experiments simultaneously reduces the time needed for the firm to investigate the benefits of changes in operating procedures. (2) With the greater number of observations, the internal validity of any conclusions is greater. Hence, geographically dispersed firms benefit more from knowledge creation via experimentation than geographically concentrated firms. Second, geographic dispersion of plants will affect the efficiency of knowledge transfer. Here the single-plant firm has an advantage – for reasons similar to those put forward by Mokyr for the development of the factory during the industrial revolution. In particular the transfer of tacit knowledge is difficult even with face-to-face contact but without it such a transfer may be nearly impossible. A second issue with geographic dispersion relates to the usefulness of the knowledge being transferred. As the similarity of the plants declines, the usefulness of information learned at one location to other locations is reduced. In short, geographically dispersed firms benefit less from firm knowledge than geographically concentrated firms.

As pointed out by Mokyr (2002, p. 141), the Demsetz/Becker-Murphy framework also suggests that when knowledge can be shared and believed among agents without the need for personal contact then firms may survive, but the large plants we know today may become less necessary. This point is becoming more important as the use of ICTs is expanding.

The development of ICTs has meant that the costs of moving people as opposed to moving information have risen sharply. The costs involved in sending and receiving information have fallen thanks to technologies such as email and the Internet along with falls in the costs of long-distance phone calls and the expanding use of cellular networks. The costs of people moving have not fallen so dramatically. Commuting to work via congested city and suburban streets, for example, is at least as difficult as it was two decades ago. The increasing interest in congestion pricing in many cities around the world suggests that traffic problems are not lessening. The increasing relative cost of moving people as opposed to information would suggest that the size of the ‘unit of production’ should be moving away from the large factory, so dominant for the last two centuries, towards more home-based production, as in the period before the industrial revolution. Mokyr (2002, p. 155) does, however, add a cautionary note. He argues that the movement away from work in the factory setting will at some point run into diminishing returns and what we will see is the locus of work remaining a mixture of work at home and work away from home.

McDermott (2001, pp. 52–3) reinforces this cautionary note by raising four issues. First, McDermott argues that monitoring remotely would be problematic given that to be effective, it may violate the norms of privacy. Second, joint production in the home of market and household goods could diminish in the future. The increased market provision of ‘household goods’ – day care, mobile dog grooming, internet shopping and home delivery, etc. – all allow workers to spend an increased amount of time away from home. Third, the growth in personal services means that in the future many workers will have to serve people directly in ways that would be difficult to do remotely (personal trainers, customer service representatives, mechanics, craftsmen, for example). Finally, to quote McDermott (2001, p. 53): ‘As Mokyr notes, citing Gavin Wright, “In the limit we could devise an economy in which technology is designed by geniuses and operated by idiots”. In that case, home production for market may indeed take off. But I am uneasy with that conclusion. It seems to me that a large part of the population may, unfortunately, substitute information and computing power for their own brain development, but these workers will not be the kind that will be left alone to telecommute. These workers will require considerable oversight in something like a factory.’

Brynjolfsson (1994, p. 1654) also sees advantages in firms being small when information is important in production. In his view, smaller firms have an advantage in providing incentives both because it is easier to separate out the contributions made by each individual, and thus to reward each individual accordingly, and because it is more likely that agents in small firms have a stronger incentive to make non-contractible contributions. Small firms therefore have an advantage over larger ones in situations in which it is important to provide incentives for the application of information in ways that cannot be easily foreseen and incorporated into a contract. Brynjolfsson (1994, footnote 12) goes further by noting that the stronger, output-based incentives for the non-contractible actions in smaller firms will not only induce higher effort overall, but, in multidimensional models, also induce less effort on actions that do not enhance output. As firms become smaller, the probability that the unit of production of a firm will coincide with the household increases.

Both the Demsetz/Becker-Murphy and Brynjolfsson models indicate that when knowledge is an important factor of production, small firms have advantages. If Mokyr is right, then this downsizing of firms should lead to a movement back towards home production and away from large factory production. But even if this is so, it is not clear whether these home producers will be single-unit firms or units of a multiunit company.

## VI THE MANAGEMENT LITERATURE

In her seminal work, Penrose (1959) outlines the principles governing the growth and diversification of firms and the rate at which firms can grow efficiently. Importantly, Penrose views the firm as encompassing a broader set of productive resources than the standard economic set of inputs – land, labour and capital. For her firms should be seen as collections of resources and services derived

from these resources, all organised under an administrative framework. But this idea received little formal attention before the 1980s. Wernerfelt (1984) was one of the first papers to formally analyse the firm from the resource side. Wernerfelt (1984, p. 172) identifies a resource as a tangible or intangible asset that is semipermanently tied to the firm. Such assets could be, for example, brand names, in-house knowledge of technology, skilled employees, efficient procedures, etc., and can be either a strength or a weakness of the firm.<sup>22</sup> Because resources mesh in a team-like way, they are worth more to the firm than they can be sold for on the market. Thus, they yield rents to the firm controlling them. Firms can act to develop resource position barriers to help protect them from incumbents in much the same way as an entry barrier protects them from potential entrants. Acquisitions and merges can be seen, in Wernerfelt's view, as ways of trading otherwise non-tradable resources. The central issues for Wernerfelt are diversification and growth of the firm.

Barney (1991) extends the strategic management literature on competitive advantage by looking at the case where resources are heterogeneously distributed across firms and these differences are stable over time. He examines the link between firm resources and sustained competitive advantage. The major issue for Barney is how ownership and control of resources can be used to achieve and sustain a competitive advantage. He sees the firm as being concerned with seeking unique, or at least hard to imitate, inputs. Primary attention is focused on how a manager can identify resources that are likely to generate rents and incorporate characteristics that results in these inputs being difficult to copy. Thus, the central activity of the firm is seen as the deployment and combination of specific inputs.

While the Barney and Wernerfelt papers broaden our understanding of firm strategy and performance they do not develop the resource-based view point into a theory of the firm *per se*.<sup>23</sup> But such a development did begin, roughly, contemporaneously with Wernerfelt (1984) and Barney (1991).

Since the early 1990s a knowledge-based (or resource-based) theory of the firm has been developing in the management literature. The major source of momentum behind this literature was the publication of papers by Conner (1991) and Kogut and Zander (1992). An important point common to all knowledge-based approaches to the theory of the firm is the rejection of the pure contractual interpretation of the nature of the firm that underlies the economic theories of the firm. In the resource-based theory, it is argued that the firm is essentially a knowledge-bearing entity and that it is possible to conceive of a theory concerning the existence, boundaries and internal organisation of the firm without reference to property rights, incentives and opportunism.

Early in their article, Kogut and Zander (1992, p. 384) state what amounts to their central proposition:

<sup>22</sup> No longer productive assets may be difficult for firms to rid themselves of.

<sup>23</sup> In an article reviewing the origins and impact of Wernerfelt (1984), Lockett *et al.* (2008, p. 1130) write: 'Wernerfelt is clear that the RBV [resource-based view] is not a theory of the firm, because it does not address issues of firm existence and boundaries'.

Our view differs radically from that of the firm as a bundle of contracts that serves to allocate efficiently property rights. In contrast to the contract approach to understanding organizations, the assumption of the selfish motives of individuals resulting in shirking or dishonesty is not a necessary premise in our argument. Rather, we suggest that organizations are social communities in which individual and social expertise is transformed into economically useful products and services by the application of a set of higher-order organizing principles. Firms exist because they provide a social community of voluntaristic action structured by organizing principles that are not reduceable to individuals.

What is important for an organisation is that a large part of the stock of knowledge is tract and social. Knowledge is produced in a social setting and is inseparable from that setting. Knowledge is accumulated through experience and is thus path dependent. Hence, each organisation is a distinct social and historical entity.

An important implication of this is that when a firm is viewed as a repository of tacit, social and path-dependent organisational knowledge, issues to do with the existence, boundaries and internal organisation of the firm are seen differently than in the contractual approach to the firm. As already noted, the firm is a bearer of specialised and tacit knowledge, which can be efficiently employed only when applied within that particular firm. This is because it is difficult, if not impossible, to take just a piece of a given firm's tacit, social and specific knowledge and apply it successfully within some other organisation. Firms exist, therefore, because they provide, in Kogut and Zander's view, a 'social community of voluntaristic action' in which this firm-specific knowledge can be produced and applied. The advantage that a firm has over the use of markets is that it can supply the required, if somewhat vaguely defined, 'higher order organizing principles'.<sup>24</sup> Such principles cannot, apparently, be supplied by the market.

Conner (1991) reaches a similar conclusion as Kogut and Zander (1992), in that she also sees that the existence of firms can be explained in a manner conceptually very different from the more standard contractual approach. On the existence issue, Conner (1991, p. 139) argues that:

For resource-based theory, existence needs to be explained in terms of a firm's superiority to two alternative forms of organization: a collection of market contracts and other firms. By the latter, the intention is to raise the issue of why a particular firm exists, as opposed to its assets being distributed among other firms.

Consideration of transaction costs, opportunism and incentives play no role in Conner's explanation; co-specialisation of assets is the important issue. To deal

<sup>24</sup> These principles would seem to include things like, 'shared coding schemes', 'a set of values' that are learned, 'a shared language' and 'a set of higher-order organizing principles [which] act as mechanisms by which to codify technologies into a language accessible to a wider circle of individuals' (Kogut and Zander, 1992, p. 389).

with the second issue first: Why does a particular firm exist in terms of its relation to other firms? Conner (1991, p. 139) argues that heterogeneous firms exist because the assets with which they work are themselves heterogeneous, making each asset a 'better fit' with some firms than with other firms.<sup>25</sup>

As to the first existence concern, consider an example from Conner (1991, pp. 141–2). Let there be two related activities,  $S$  and  $T$ ; then the question is, when are  $S$  and  $T$  best owned in common rather than being owned independently and having a contractual relationship between them? The notation  $ST$  will be used to represent the case where  $S$  and  $T$  are owned together, which is a project involving both activities undertaken 'in house'. When the notation  $S+T$  is used, it refers to the case where the project is undertaken but ownership of  $S$  and  $T$  is separated and the project involves a contract over the market interface. An important point here is that, as separate firms,  $S$  and  $T$  will have their own CEO and management along with their own set of firm-level routines, cultures, etc. Then, what is the difference between  $ST$  and  $S+T$ ? In Conner's view,  $S+T$ , as compared with  $ST$ , should find it more difficult to organise the research in such a way as to make the knowledge and skills accessible to both  $S$  and  $T$  as separate operations. This is simply because as Conner (1991, p. 142) explains it, '...  $S+T$  must try to orient in two different directions, one for  $S$  and another for  $T$ . That is, any gains from the relationship, new knowledge, skills and routines, etc., need to be integrated into two firm-specific cultures rather than just one. Cooperation in terms of just  $ST$  is easier because it is needed in only one direction. The implication of this is that these gains from cooperation and coordination associated with  $ST$ , relative to  $S+T$ , are sufficient to explain the existence of  $ST$ . The general issue is formulated by Conner as:

To encapsulate the argument to follow, the case made here is that, under certain circumstances, firms have advantage over market relationships in the joint activity of creating and redeploying specific capital. Further, the advantage of firms in the creation-redeployment combination need not stem from an opportunism-control advantage. Instead, it is argued that the firm has advantage over a collection of market transactions in those situations where redeployment inside the firm is more efficient and, perhaps more important, qualitatively more productive because of the opportunity to benefit from asset interdependencies within the firm.

(Conner, 1991, p. 142)

Thus, the resource-based or knowledge-based theory of the firms claims to be able to explain the existence (and most probably the boundaries and internal organisation) of the firm without reference to ideas like property rights, incentives and opportunism, which underlie the orthodox theories of the firm discussed in the previous sections. This claim has not gone unchallenged. Foss (1996a) offers one of the most powerful critiques of this approach.

<sup>25</sup> Firms are seen as having heterogeneous asset bases in that each has a unique bundle of resources and relationships.

Foss (1996a, p. 473) argues that both Conner's and Kogut and Zander's approaches are versions of 'technological determinism'.<sup>26</sup> That is, the idea that technology directly determines economic organisation. In Foss's view, Conner and Kogut and Zander fall prey to technological determinism when they argue that the need for things like shared codes, languages, etc., necessitates firm organisation in a manner independent of more standard ideas like opportunism and asymmetric information.

Recall that, in the Conner/Kogut and Zander story, to fully utilise assets and resources, these assets and resources have to be embedded in higher order organising principles. But such embeddedness does not, in the Foss argument, presuppose common ownership. Separately owned activities could be conceptually more 'embedded' in this sense than, as an example, divisions of the same firm. Higher order organising principles are not a distinguishing feature of firms relative to markets. Such principles may be more abundant in firms than markets, but they exist in both. In fact, Foss (1996a, p. 473) goes further and argues that for firms to cultivate better higher order organising principles requires not moral utopia, but opportunism. Consider a state of moral utopia, characterised by the total absence of opportunism. In such a world any gains from assets being embedded in higher order organising principles can be achieved over the market. Economic agents (human resources) could simply meet at a given location, own their own physical capital (or rent it from each other) and develop, as a team, value-enhancing higher order organising principles among themselves. Given the absence of opportunism, Foss argues, the level of co-specialisation of resources carries no implications for ownership. In terms of the Conner example above, *ST* and *S+T* would do just as well as each other.

In a response to Foss (1996a), Conner and Prahalad (1996) argue that an independent-of-opportunism rationale for the organisation of firms can be developed. They argue that, relative to independent contractors, firms can economise on a number of opportunism-independent transaction costs. These costs relate to flexibility, communication, learning, bargaining, and possibly also measurement. Conner and Prahalad see the rationale for the firm in terms of its superior coordinating ability including its capacity to reduce costs associated with information and bargaining. The crucial concepts in Conner and Prahalad's argument are the 'knowledge-substitution effect' and the 'flexibility effect'. Put simply, the knowledge-substitution effect means that when an employee takes direction from a manager, the manager's knowledge is a partial substitute for the employee's knowledge, while the flexibility effect is due to the fact that an employment contract brings greater flexibility, relative to market contracting, since the employment contract does not require renegotiation. In the Conner and Prahalad view these effects are sensitive to organisational modes. To the degree that firms do better in advancing these two effects, this provides an opportunism-independent rationale for the existence of the firm.

<sup>26</sup>See Williamson (1985, pp. 86–90).

Foss (1996b, pp. 520–1) replies that the Conner and Prahalad view that the employment contract is sufficient to characterise the firm is incomplete. He argues that the ownership of physical assets is an indispensable element for any understanding of the firm and that the allocation of control rights to these assets will reflect, in part, the opportunity for opportunism. Foss also argues that both the flexibility and knowledge-substitution effects can be achieved through market contracting.<sup>27</sup>

Nickerson and Zenger (2004, pp. 617–8) develop three additional criticisms of the knowledge-based literature. First, they argue that two common and fundamental arguments within it are contradictory. One group argues that hierarchies exist to avoid knowledge transfer. The emphasis is on the firm's ability to direct, via authority, the activities of others. A second group view hierarchies as enabling knowledge transfer. They emphasis the firm's ability to support the formation of shared identity and language. Second, Nickerson and Zenger claim that the literature has not provided a theory that predicts when hierarchies are preferred to markets or vice versa. Third, they argue that the literature focuses on knowledge exchange rather than knowledge production.

Hence, Nickerson and Zenger (2004, p. 618) conclude, '[i]n summary, we do not yet have a knowledge-based theory of the firm'. While Foss (1996a, p. 473) says, '... co-specialization and the presence of higher order organizing principles are not sufficient to explain the existence of the specific constellation of property rights that characterizes the firm'. It is therefore not clear how this approach can explain the changes to the firm we see happening as a result of the growth of the knowledge economy.

Nickerson and Zenger (2004) is an attempt to address some of these shortcomings. They set out to develop a theory that explains how the prospective objectives for knowledge generation determine the choice of organisational structure. For Nickerson and Zenger, the fundamental question is to do with how a manager should organise individuals in pursuit of the knowledge the firm desires. They wish to explain how different organisational structures influence the efficient production of knowledge valuable to a firm. To this end they start by defining the unit of analysis for knowledge production to be a specific problem.<sup>28</sup> The value of this problem is determined by both the values in the array of possible solutions and the costs of discovering a valuable problem. They assume that the solution to complex problems is a unique combination of existing knowledge. An important distinction Nickerson and Zenger make is between decomposable and non-decomposable problems.

Decomposable problems involve limited interaction between agents attempting to solve them, while non-decomposable problems require much greater interaction. This has important implications for the type of searching used to find a solution for a given problem. There are two general methods of searching. First, there is directional search, which amounts to trial and error searching.

<sup>27</sup> In addition to Conner and Prahalad (1996), Kogut and Zander (1996) also responded to Foss (1996a). Foss (1996b) is a reply to both these papers.

<sup>28</sup> A valuable problem is one which if solved will yield desirable knowledge.

This is effective only for the case of decomposable problems. The second method is heuristic search. This refers to the situation where a team cognitively evaluates a range of probable consequences of design choices. Non-decomposable problems require specialised knowledge to be shared. Such sharing is impeded by two barriers: (1) people are limited in the speed with which they can learn and (2) they are subject to self-interest. The interaction between distributed knowledge and self-interest gives rise to two hazards for the exchange of knowledge: (1) knowledge appropriation and (2) strategic knowledge accumulation. As a consequence, efficiency considerations require the selection of an optimal governance structure and the provision of incentives.

Nickerson and Zenger examine three distinct governance mechanisms and evaluate their suitability for dealing with problems of different types. The governance mechanisms are (1) markets, (2) authority-based hierarchies and (3) consensus-based hierarchies. They conclude that markets are well suited for situations in which the problem is decomposable and directional search is thought appropriate; that consensus-based hierarchies generate large organisational costs and thus should be used only when the gains from consensus are large, which is the case for problems that are highly complex and non-decomposable; and that an authority-based hierarchy is preferable to the use of markets for cases where heuristic search is applicable and not preferred for those situations where directional search is optimal. The use of authority-based hierarchy is thought best when problems are moderately complex.<sup>29</sup>

Foss (2007, p. 44) argues that the use of the problem as the unit of analysis is limited in that it is, in the main, designed to help an understanding of the governance of knowledge creation – that is, problem solving by combining knowledge – not an understanding of knowledge sharing or integration. An additional point that is not clear from the discussion above is how ‘new economy’ firms differ from ‘old economy’ firms. Both old and new economy firms have problems to solve and can choose between the alternatives Nickerson and Zenger consider to solve them. Thus, it is not clear how this approach can explain the changes to the firm we see happening as a result of the knowledge economy.

## VII CONCLUSION

Current research offers an increased understanding of how markets operate in the knowledge economy and some understanding of the effect of this on corporate strategy and related management issues.<sup>30</sup> This, however, stops short of an actual theory of the knowledge firm. While it is, as Foss (1997, p. 309) notes, ‘... generally recognized that knowledge problems are behind all interesting aspects of economic organization, and that the major function of, for example, firms is to cope with the economic problems introduced by

<sup>29</sup> Foss (2007) gives an much more expansive discussion of the whole ‘knowledge governance approach’ including the Nickerson and Zenger (2004) contribution.

<sup>30</sup> See, for example, Arora *et al.* (2001), Carley (2000), Kling and Lamb (2000), Leibold *et al.* (2005), Shapiro and Varian (1999) and Smith *et al.* (2000).

changing, partial, tacit, complex, asymmetrical, etc., knowledge', it must also be recognised that none of the orthodox theories of the firm captures the full implications of knowledge for economic organisation.

The previous sections made this point by briefly outlining the effects of the increasing importance of knowledge for the mainstream theories of the firm. Relationship-specific investment-induced hold-up arguments for vertical integration are at their weakest when dealing with human capital. Human capital cannot be owned in the same way as physical capital and so the investor in human capital can act opportunistically whether an employee or not. The incentive-system theory assumes the use of a physical asset rather than a human capital asset in the production process. Neither the ownership nor the value of a human asset can be transferred and so such an asset cannot determine where the boundaries of the firm lie within the model. The extensions of the GHM framework offered by Brynjolfsson (1994) and Rabin (1993) inherits the implicit owner–manager restriction of the original GHM framework and thus are of limited use when modelling the knowledge firm. When we turn to the location of production, the models suggest that we should, in general terms, see a movement back towards home production but we are not given a specific relationship between knowledge and plant size or production location. The knowledge-based theory of the firm deals no better with the increasing importance of knowledge in the economy.

We are left with an unsatisfactory model of the (knowledge) firm and thus we are unable to give guidance on either empirical or policy questions that flow, via changes to the firm, from the development of the knowledge economy. Firms' organisational structures are changing in response to the increased prominence of information and knowledge in the production process. In the new economy, not only will we see changes in the location of production, but even if production still takes place within a traditional firm, a factory or an office, that firm may have a very different structure and organisation from that which we see today. As an example, consider the case noted in Section I where Rajan and Zingales (2003, p. 87) argue that we are in fact seeing a new 'kinder, gentler firm'. This is in response to the increase in the importance of human capital, along with increased competition and access to finance, all of which have increased the worker's importance and improved the outside options for workers, thereby changing the balance of power within firms. Rajan and Zingales (2003, p. 87) also argue that the biggest challenge for the owners and management today is to manage in an environment of much reduced authority. Today, authority has to be gained by persuading workers that the workplace is an attractive one and one that they would hate to lose. To do this, management has to ensure that work is enriching, that responsibilities are handed down, and that rich bonds develop not only among workers but also between workers and management.

Another example is Cowen and Parker (1997) who make a similar point about the changes to organisational structures that the new economy is bringing about. For them:

Information as a factor of production is making old functional structures and methods of organisation and planning redundant in many areas of business.

The successful use of knowledge involves not only its generation, but also its mobilisation and integration, requiring a change in the way it is handled and processed.

(Cowen and Parker, 1997, p. 12)

Organisational change, according to Cowen and Parker, is the consequence of the increasing need to make use of market principles within the firm and the growing importance of human capital. They note that as far as a firm's labour force is concerned, the emphasis has now shifted towards encouraging knowledge acquisition, skills and adaptability since these are seen as critical in maintaining a competitive advantage. (Cowen and Parker, 1997, p. 32).

Firms are obliged to rely more on market-based mechanisms as the most efficient way of processing and transmitting information and giving the firm the flexibility but also the focus it requires. Companies are decentralising their management systems as a way of coping with the uncertainty and pace of change in their markets. The aim is to ensure that those with the required knowledge and right incentives are the ones making the decisions and taking responsibility for the outcomes. Cowen and Parker (1997, pp. 25–8) emphasise how advances in ICTs underlie the capacity to combine the advantages of this organisational flexibility with mass production.<sup>31</sup>

Another pressure on the organisation of firms is suggested by Klein (1988). Insofar as he is right that hold-up problems can be based on human capital as much as on physical capital, it is a problem that will, most likely, only get worse as we move towards more human capital-based firms. Klein argues that vertical integration can solve potential hold-up problems even if they hinge on human capital. He sees vertical integration as organisational ownership. The major reason human capital-based hold-up cannot occur after integration is because collusion is difficult with a large number of agents. As Klein (1988, p. 220) explains in terms of the General Motors Fisher Body example, post-integration, the Fisher brothers could not hold up General Motors by telling all the employees to leave General Motors and turn up at a new address on Monday morning. This is neither legally possible, since the workers are no longer employees of Fisher Body, nor economically feasible. But if integration is, at least partly, an answer to human capital hold-up, then it is a reason for firms to become larger rather than smaller as has been argued above.<sup>32</sup>

<sup>31</sup> In addition to Cowen and Parker (1997), see Gable and Ellig (1993) and Koch (2007) for more detail on market-based management.

<sup>32</sup> Klein's argument is somewhat problematic. Freeland (2000) raises one issue in that he sees vertical integration creating, not solving, hold-up problems. Freeland argues that vertical integration can increase vulnerability to hold-up as it did for General Motors. He argues that in the case of the General Motors Fisher Body integration, hold-up was not an issue before integration but that the Fisher brothers successfully held up General Motors post-integration. Thus far from reducing opportunistic behaviour, vertical integration increased General Motors's vulnerability to rent-seeking behaviour based on human asset specificity. In addition, the Klein argument is a 'large numbers argument'. If knowledge-based firms are small for other reasons, the Klein argument loses much of its power. With small numbers collusion is much easier.

Another issue with knowledge firms is what would bankruptcy mean for them? If the firm comprises only human capital resources, whose accumulated knowledge is the source of wealth creation, then predictions about what would happen at the dissolution of the firm are unclear. Who has the rights to the sell-off of the assets, when these assets are embodied in human beings? How can these assets be sold-off? How could the 'shareholders' or debt holders recover any of their investments?

But little of these types of changes, trade-offs and issues are captured or explained by the mainstream theory of the firm. Expanding the orthodox view of the firm to include the new reality of the knowledge economy should be an urgent issue on the economic research agenda. As was argued in Section I, changes to the firm matter simply because so much economic activity takes place within their boundaries. As a consequence, changes to the firm will help determine who are the 'winners and losers' from the knowledge economy. As in all previous 'economic revolutions', this is the ultimate issue with the knowledge economy.

#### ACKNOWLEDGEMENT

I wish to thank, without implicating, an anonymous referee, Philip Gunby, Bob Hamilton, Andrea Menclova, Les Oxley and Glenda Park.

#### REFERENCES

- ALCALY, R. (2003). *The New Economy: What it is, How it Happened, and Why it is Likely to Last*. New York: Farrar, Straus and Giroux.
- ALONSO, R., DESSEIN, W. and MATOUSCHEK, N. (2008). When does coordination require centralization? *American Economic Review*, **98**, 1, pp. 145–79.
- ARORA, A., FOSFURI, A. and GAMBARDELLA, A. (2001). *Markets for Technology: The Economics of Innovation and Corporate Strategy*. Cambridge, MA: MIT Press.
- AUDIA, P. G., SORENSON, O. and HAGE, J. (2001). Tradeoffs in the organization of production: multiunit firms, geographic dispersion and organizational learning. In J. A. C. Baum and H. R. Greve (eds.), *Multiunit Organization and Multimarket Strategy: Advances in Strategic Management*, Vol. 18. Oxford: JAI Press, pp. 75–105.
- BABBAGE, C. (1835). *On the Economy of Machinery and Manufactures* (4th edn, enlarged). London: C. Knight.
- BAIRD, D. G. (2003). In Coase's Footsteps. John M. Olin Law & Economics Working Paper No 175 (2nd series). The Law School, the University of Chicago, January.
- BARNEY, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, **17**, 1, pp. 99–120.
- BECKER, G. S. and MURPHY, K. M. (1992). The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics*, **107**, 4, pp. 1137–60.
- BOLTON, P. and DEWATRIPONT, M. (1994). The firm as a communication network. *Quarterly Journal of Economics*, **109**, 4, pp. 809–40.
- BOLTON, P. and DEWATRIPONT, M. (2005). *Contract Theory*. Cambridge, MA: MIT Press.
- BRYNJOLFSSON, E. (1994). Information assets, technology, and organization. *Management Science*, **40**, 12, pp. 1645–62.
- BRYNJOLFSSON, E. and KAHIN, B. (eds.), (2000). *Understanding the Digital Economy: Data, Tools, and Research*. Cambridge, MA: MIT Press.
- BUTLER, E. (2007). *Adam Smith – A Primer*. London: Institute of Economic Affairs.
- CARLEY, K. (2000). Organizational change and the digital economy: a computational organization science perspective. In E. Brynjolfsson and B. Kahin (eds.), *Understanding the Digital Economy: Data, Tools, and Research*. Cambridge, MA: MIT Press, pp. 325–51.

- CASADESUS-MASANELL, R. and SPULBER, D. F. (2000). The fable of Fisher Body. *Journal of Law and Economics*, **43**, 1, pp. 67–104.
- COASE, R. H. (1937). The nature of the firm. *Economica*, **4**, 16, pp. 386–405.
- COASE, R. H. (1992). The institutional structure of production. *The American Economic Review*, **82**, 4, pp. 713–9.
- COASE, R. H. (2000). The acquisition of Fisher Body by General Motors. *Journal of Law and Economics*, **43**, 1, pp. 15–31.
- COASE, R. H. (2006). The conduct of economics: the example of Fisher Body and General Motors. *Journal of Economics and Management Strategy*, **15**, 2, pp. 255–78.
- 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**, 1, pp. 121–54.
- CONNER, K. R. and PRAHALAD, C. K. (1996). A resource-based theory of the firm: knowledge versus opportunism. *Organization Science*, **7**, 5, pp. 477–501.
- COWEN, T. and PARKER, D. (1997). *Markets in the Firm: A Market-Process Approach to Management*. London: Institute of Economic Affairs.
- COYLE, D. (2001). *Paradoxes of Prosperity: Why the New Capitalism Benefits All*. London: Texere.
- CREMER, J., GARICANO, L. and PRAT, A. (2007). Language and the theory of the firm. *Quarterly Journal of Economics*, **122**, 1, pp. 373–407.
- DEBREU, G. (1959). *Theory of Value*. New York: Wiley.
- DE MEZA, D. and LOCKWOOD, B. (1998). Does asset ownership always motivate managers? Outside options and the property rights theory of the firm. *Quarterly Journal of Economics*, **113**, 2, pp. 361–86.
- DEMSETZ, H. (1982). *Economic, Legal, and Political Dimensions of Competition*. Amsterdam: North-Holland Publishing Company.
- DEMSETZ, H. (1988). The theory of the firm revisited. *Journal of Law, Economics, and Organization*, **4**, 1, pp. 141–61.
- DEMSETZ, H. (1995). *The Economics of the Business Firm: Seven Critical Commentaries*. Cambridge: Cambridge University Press.
- DESSEIN, W. and SANTOS, T. (2003). Demand for coordination. National Bureau of Economic Research (NBER) Working Paper No. 10056, November.
- FOSS, N. J. (1996a). Knowledge-based approaches to the theory of the firm: some critical comments. *Organization Science*, **7**, 5, pp. 470–6.
- FOSS, N. J. (1996b). More critical comments on knowledge-based theories of the firm. *Organization Science*, **7**, 5, pp. 519–23.
- FOSS, N. J. (1997). The classical theory of production and the capabilities view of the firm. *Journal of Economic Studies*, **24**, 4–5, pp. 307–23.
- FOSS, N. J. (2007). The emerging knowledge governance approach: challenges and characteristics. *Organization*, **14**, 1, pp. 29–52.
- FOSS, N. J., LANDO, H. and THOMSEN, S. (2000). The theory of the firm. In B. Bouckaert and G. De Geest (eds.), *Encyclopedia of Law and Economics*, Vol. III. Cheltenham, UK: Edward Elgar Publishing Ltd, pp. 631–58.
- FREELAND, R. F. (2000). Creating holdup through vertical integration: Fisher Body revisited. *Journal of Law and Economics*, **43**, 1, pp. 33–66.
- GABLE, W. and ELLIG, J. (1993). *Introduction to Market-Based Management*. Fairfax: Center for Market Processes.
- GALOR, O. and MOAV, O. (2006). Das human-kapital: a theory of the demise of the class structure. *Review of Economic Studies*, **73**, 1, pp. 85–117.
- GARICANO, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, **108**, 5, pp. 874–904.
- GIBBONS, R. (2005). Four formal(izable) theories of the firm? *Journal of Economic Behavior and Organization*, **58**, 2, pp. 200–45.
- GOLDBERG, V. P. (2008). Lawyers asleep at the wheel? The GM–Fisher Body contract. *Industrial and Corporate Change*, **17**, 5, pp. 1071–84.
- GROSSMAN, S. J. and HART, O. D. (1986). The costs and benefits of ownership: a theory of vertical and lateral integration. *Journal of Political Economy*, **94**, 4, pp. 691–719.
- GROSSMAN, S. J. and HART, O. D. (1987). Vertical integration and the distribution of property rights. In A. Razin and E. Sadka (eds.), *Economic policy in theory and practice*. London: Macmillan Press, pp. 504–46.

- HART, O. D. (1988). Incomplete contracts and the theory of the firm. *Journal of Law, Economics, and Organization*, **4**, 1, pp. 119–39.
- HART, O. D. (1989). An economist's perspective on the theory of the firm. *Columbia Law Review*, **89**, 7, pp. 1757–74.
- HART, O. D. (1995). *Firms, Contracts, and Financial Structure*. Oxford: Oxford University Press.
- HART, O. D. (2007). Hold-up, asset ownership, and reference points. National Bureau of Economic Research (NBER) Working Paper No. 13540, October.
- HART, O. D. and HOLMSTRÖM, B. (2008). A theory of firm scope. National Bureau of Economic Research (NBER) Working Paper No. 14613, December.
- HART, O. D. and MOORE, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, **98**, 6, pp. 1119–58.
- HART, O. D. and MOORE, J. (2007). Incomplete contracts and ownership: some new thoughts. *American Economic Review*, **97**, 2, pp. 182–6.
- HOLMSTRÖM, B. (1999). The firm as a subeconomy. *Journal of Law, Economics, and Organization*, **15**, 1, pp. 74–102.
- HOLMSTRÖM, B. and MILGROM, P. (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, **7**, (Special Issue), pp. 24–52.
- HOLMSTRÖM, B. and MILGROM, P. (1994). The firm as an incentive system. *American Economic Review*, **84**, 4, pp. 972–91.
- HOLMSTRÖM, B. and TIROLE, J. (1991). Transfer pricing and organizational form. *Journal of Law, Economics, and Organization*, **7**, 2, pp. 201–28.
- KAY, J. (2005). *Culture and Prosperity: Why Some Nations are Rich but Most Remain Poor*. New York: Harper Business.
- KENNEDY, G. (2005). *Adam Smith's Lost Legacy*. Houndmills Basingstoke Hampshire: Palgrave Macmillan.
- KENNEDY, G. (2008). *Adam Smith: A Moral Philosopher and His Political Economy*. Houndmills Basingstoke Hampshire: Palgrave Macmillan.
- KIM, S. (1999). The rise of multiunit firms in U.S. manufacturing. *Explorations in Economic History*, **36**, 4, pp. 360–86.
- KIM, S. (2001). Markets and multiunit firms from an American historical perspective. In J. A. C. Baum and H. R. Greve (eds.), *Multiunit Organization and Multimarket Strategy: Advances in Strategic Management*, Vol. 18. Oxford: JAI Press.
- KLEIN, B. (1988). Vertical integration as organizational ownership: the Fisher Body–General Motors relationship revisited. *Journal of Law, Economics, and Organization*, **4**, 1, pp. 199–213.
- KLEIN, B. (1996). Why hold-ups occur: the self-enforcing range of contractual relationships. *Economic Inquiry*, **XXXIV**, pp. 444–63.
- KLEIN, B. (2000). Fisher–General Motors and the nature of the firm. *Journal of Law and Economics*, **43**, 1, pp. 105–41.
- KLEIN, B. (2007). The economic lessons of Fisher Body–General Motors. *International Journal of the Economics of Business*, **14**, 1, pp. 1–36.
- KLEIN, B. (2008). The enforceability of the GM–Fisher Body contract: comment on Goldberg. *Industrial and Corporate Change*, **17**, 5, pp. 1085–96.
- KLEIN, B., CRAWFORD, R. G. and ALCHIAN, A. A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics*, **21**, 2, pp. 297–326.
- KLING, R. and LAMB, R. (2000). IT and organizational change in digital economies: a sociotechnical approach. In E. Brynjolfsson and B. Kahin (eds.), *Understanding the Digital Economy: Data, Tools, and Research*. Cambridge Mass.: MIT Press.
- KNIGHT, F. H. (1921). *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin Company.
- KOCH, C. G. (2007). *The Science of Success: How Market-Based Management Built the World's Largest Private Company*. Hoboken NJ: John Wiley & Sons, Inc.
- KOGUT, B. and ZANDER, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, **3**, 3, pp. 383–97.
- KOGUT, B. and ZANDER, U. (1996). What firms do? Coordination, identity, and learning. *Organization Science*, **7**, 5, pp. 502–18.
- LAMOREAUX, N. R., RAFF, D. M. G. and TEMIN, P. (2003). Beyond markets and hierarchies: toward a new synthesis of American business history. *American Historical Review*, **108**, 2, pp. 404–33.

- LEIBOLD, M., PROBST, G. J. B. and GIBBERT, M. (2005). *Strategic Management in the Knowledge Economy: New Approaches and Business Applications* (2nd-updated edition). Weinheim: Wiley-VCH.
- LOCKETT, Andy, O'SHEA, R. P. and WRIGHT, M. (2008). The development of the resource-based view: reflections from Birger Wernerfelt. *Organization Studies*, **29**, 8–9, pp. 1125–41.
- MCDERMOTT, J. (2001). The rise and fall of the factory system: technology, firms, and households since the industrial revolution: A comment. *Carnegie-Rochester Conference Series on Public Policy*, **55**, pp. 47–54.
- MCMILLAN, J. (2002). *Reinventing the Bazaar: A Natural history of Markets*. New York: W.W. Norton and Company.
- MOKYR, J. (2001). The rise and fall of the factory system: technology, firms, and households since the industrial revolution. *Carnegie-Rochester Conference Series on Public Policy*, **55**, pp. 1–45.
- MOKYR, J. (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton: Princeton University Press.
- MOORE, J. (1992). The firm as a collection of assets. *European Economic Review*, **36**, 2–3, pp. 493–507.
- NICKERSON, J. A. and ZENGER, T. R. (2004). A knowledge-based theory of the firm: the problem-solving perspective. *Organization Science*, **15**, 6, pp. 617–32.
- OXLEY, L., WALKER, P., THORNS, D. and WANG, H. (2008). The knowledge economy/society: the latest example of “Measurement without theory”? *The Journal of Philosophical Economics*, **2**, 1, pp. 20–54.
- PENROSE, E. (1959). *The Theory of the Growth of the Firm*. Oxford: Basil Blackwell.
- RABIN, M. (1993). Information and the control of productive assets. *Journal of Law, Economics, and Organization*, **9**, 1, pp. 51–76.
- RAJAN, R. G. and ZINGALES, L. (1998). Power in the theory of the firm. *Quarterly Journal of Economics*, **113**, 2, pp. 387–432.
- RAJAN, R. G. and ZINGALES, L. (2003). *Saving Capitalism From the Capitalists: Unleashing the Power of Financial Markets to Create Wealth and Spread Opportunity*. New York: Crown Business.
- ROBERTS, J. (2004). *The Modern Firm: Organizational Design for Performance and Growth*. Oxford: Oxford University Press.
- ROSENBERG, N. (1994). *Exploring the Black Box: Technology, Economics, and History*. Cambridge: Cambridge University Press.
- SCHERER, F. M., BECKENSTEIN, A., KAUFER, E. and MURPHY, R. D. (1975). *The Economics of Multi-plant Operation: An International Comparison Study*. Cambridge MA: Harvard University Press.
- SHAPIRO, C. and VARIAN, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press.
- SIMON, H. (1991). Organizations and markets. *Journal of Economic Perspectives*, **5**, 2, pp. 25–44.
- SMITH, M. D., BAILEY, J. and BRYNJOLFSSON, E. (2000). Understanding digital markets: review and assessment. In E. Brynjolfsson and B. Kahin (eds.), *Understanding the Digital Economy: Data, Tools, and Research*. Cambridge Mass.: MIT Press.
- SFULBER, D. F. (2008). ‘Discovering the Role of the Firm: The Separation Criterion and Corporate Law’, *Northwestern Law & Economics Research Paper No. 08-23*, December 6.
- WERNERFELT, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, **5**, 2, pp. 171–80.
- WERNERFELT, B. (1997). On the nature and scope of the firm: an adjustment-cost theory. *Journal of Business*, **70**, 4, pp. 489–514.
- WILLIAMSON, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: The Free Press.
- WILLIAMSON, O. E. (1985). *The Economic Institutions of Capitalism*. New York: The Free Press.
- WILLIAMSON, O. E. (1996). *The Mechanisms of Governance*. Oxford: Oxford University Press.

Date of receipt of final manuscript: 30 January 2009