

(Final version)

On the Opportunism-Independent Theory of the Firm

James H Love
Aston Business School
Aston University
Birmingham B4 7ET
United Kingdom

Tel: 0121 359 3611
Fax: 0121 333 3474
E-mail: j.h.love@aston.ac.uk

Acknowledgements I am grateful for the constructive comments of John Finch and an anonymous referee.

On the Opportunism-Independent Theory of the Firm

Abstract

Previous developments in the opportunism-independent theory of the firm are either restricted to special cases or are derived from the capabilities or resource-based perspective. However, a more general opportunism-independent approach can be developed, based on the work of Demsetz and Coase, which is nevertheless contractual in nature. This depends on ‘direction’, that is deriving economic value by permitting one set of actors to direct the activities of another, and of non-human factors of production. Direction helps to explain not only firm boundaries and organization, but also the existence of firms without appealing to opportunism or moral hazard. I also consider the extent to which it is meaningful to speak of ‘contractual’ theories in the absence of opportunism, and whether this analysis can be extended beyond the employment contract to encompass ownership of assets by the firm.

Key words: Opportunism, Direction, Specialization, Contractual, Theory of the Firm

JEL classification: L22, L23, D23

1. Introduction

The purpose of this paper is to explore some steps in the development of an opportunism-independent theory of firm boundaries and firm existence which nevertheless retains a contractual vision of the firm. The importance of this topic lies in recent debates on the role of opportunism within two of the principal theories employed in economics and strategy: transaction-cost analysis, and the capabilities or resource-based perspective of the firm.

The centrality of opportunism to transaction-cost analysis has been the matter of debate for several years (Alchian and Woodward 1988; Williamson 1993; Stephen 1996). More recently attention has shifted to the role of opportunism within theories of the firm that are not dependent on transaction costs; predominantly, this refers to the capabilities or resource-based approach. Some writers explicitly view the resource-based approach as a form of opportunism-independent theory of the firm (Conner and Prahalad 1996). Others argue that although knowledge-based theories can tell us something about the boundaries or organization of the (multi-person) firm, they cannot give a convincing explanation for its *existence* (Foss 1996a,b), and as a result, we are not really any closer to developing an opportunism-independent theory of the existence of the firm. I suggest a different perspective, one in which an opportunism-independent vision of the firm is not necessarily linked to the capabilities approach. I argue that it is possible to say something about why firms exist in the absence of concepts such as opportunism or moral hazard, and that it is possible to have a theory of the firm which is opportunism-free, but which retains a contractual focus. This analysis depends crucially on the argument that contractual issues do not everywhere and always hinge on (Williamsonian) opportunism, and that contracts can have functions beyond those merely of incentive alignment to prevent wrongdoing.

This article proceeds as follows. First, I suggest why an opportunism-independent theory of the firm is a topic worthy of consideration, and discuss two partial developments towards such a theory. Secondly, I argue that it is possible to have opportunism-independent theories both of the existence and of the organization of the firm, and I give

examples of such theories based on the work of Harold Demsetz and Ronald Coase which depend on knowledge and information costs and specialization benefits. Thirdly, I tackle the issue of whether a contractual approach is meaningful in the absence of opportunism, and consider whether a worthwhile theory of the firm must move beyond the employment contract and deal satisfactorily with wider issues such as the ownership of assets. Finally, I outline areas in which theoretical and empirical research could help in our understanding of opportunism-independence, and suggest that, although it is indeed possible to have an opportunism-independent yet contractual vision of the firm, there is merit in other theoretical approaches in reminding us that the firm is not solely a contractual entity.

2. Opportunism-independent Approaches to the Firm

Why do we need an opportunism-independent theory?

The reason we should be interested in an opportunism-independent theory is because in its absence both economists and writers in strategy have developed an unnecessarily narrow conception both of the firm and of the determinants of its boundaries. In the case of strict adherents to transaction-cost economics, this is because of the Williamsonian view that hold-up cannot exist in the absence of opportunism. This in turn has led to an unwillingness to acknowledge hold-up problems arising from small numbers conditions which can be independent of rent-seeking opportunism or specific assets; for example, time critical production where a delay in obtaining a crucial component may give rise to such large losses that in-house production will be favoured regardless of the number of potential suppliers or the likelihood of opportunistic hold-up being exercised¹. Breaking this link between specific assets, opportunism and hold-up, unthinkable in Williamsonian transaction-cost analysis, can enrich our understanding of firm boundaries and existence. In the case of those who take a broader view of the existence of the firm, but nevertheless believe it must be inherently contractual, opportunism is central to the contractual view of

¹ Masten *et al* (1991) employ the term 'temporal specificity' to deal with the situation of time-critical production, explicitly regarding it as an element of asset specificity. In fact, time-criticality need have

the firm: what role do contracts play if parties are never at risk from opportunism? However, questioning the role of opportunism may help to bring the issue of the contractual nature of the firm into sharper focus: as I argue below, contracts can play a role which goes beyond the narrow scope of incentive alignment to protect from wrongdoing by one or other party to a contract. And finally, for those concerned with strategy, moving towards an opportunism-independent theory permits a broader conception of the firm which fits more naturally with the capabilities or resource-based view. Conner and Prahalad (1996:480) argue strongly that strategy writers need a theory of the firm to consider properly issues relating to differences in firms' attributes and performance, and Madhok (1996) sees the knowledge-based theory explicitly as "an attempt to develop an opportunism-independent theoretical understanding of the firm." (1996: 578) Therefore from several perspectives we have an interest in seeing whether it is possible to develop such a theory.

But this requires us first to be clear on what we understand by 'the firm' and by 'opportunism'. Let us accept that the firm is most satisfactorily conceived of as a multi-person entity² and ought surely to involve more than merely an employment contract, which is the limited view of the firm envisaged by *inter alia* Conner and Prahalad (1996) and Wernerfelt (1997). On opportunism, the starting point is often taken as Oliver Williamson's definition as "...self-interest seeking with guile. This includes but is scarcely limited to more blatant forms, such as lying stealing and cheating. Opportunism more often involves more subtle forms of deceit...." (Williamson, 1985: 47). Williamson does not assume that all economic agents are always opportunistic, or even that opportunism always pays. The difficulty arises in determining *ex ante* which possible contracting partners are likely to behave in this way, and discriminating between those circumstances under which opportunism will pay and those when it will not. Transaction costs can then arise from the need to protect against the likelihood of, and potential loss that may arise from, opportunistic behaviour by another party.

little to do with specific assets of any kind, nor, as suggested above, with the presence of opportunism (Stephen and Love 2000).

² There is a sub-literature on whether it is meaningful to speak of single-person entities as firms (e.g. Fourie 1993; Hodgson 1993), but for present purposes this is a somewhat sterile debate.

However, other writers have adopted visions of opportunism which do not depend on guile or deceitfulness, but which can encompass a much wide range of human behaviour such as ‘honest disagreements’ (Alchian and Woodward, 1988) and simply the act of following self-interest at the expense of others (Goldberg, 1984). As Stephen (1996) points out, this milder version of opportunism may still give rise to transaction costs because of the need for parties investing in transaction-specific capital to guard against *ex-post* hold up. This is because virtually any kind of self-interested behaviour short of actual deceit may preclude the use of a self-enforcing ‘general clause’ to get round the problem of contractual incompleteness. For present purposes, my argument is with the Williamsonian version of opportunism and its emphasis on “calculated efforts to mislead, distort, disguise, obfuscate, or otherwise obscure”. (Williamson, 1985: 47). I have no problem accepting that parties may act in a self-interested manner which can be to the detriment of others, but propose that we consider a world in which there is no need to take actions to defend against the potential returns from systematically guileful behaviour of the type described by Williamson.³ This is the sense in which I propose to consider the opportunism-independent, yet contractual, theory of the firm.

There are two strands of literature in which conceptions of the firm or its boundaries are independent of the existence of (Williamsonian) opportunism. Neither of these represents a complete theory of the firm, but they do represent stepping-stones on the way towards a more complete picture of the firm.

Divergent Capabilities

The first area is where firm boundaries may be partly determined for non-opportunistic reasons under circumstances in which there is strong divergence between the capabilities/capacities of firms, and therefore in which knowledge transfer is either impossible or prohibitively costly. This aspect is perhaps most clearly developed in theoretical writings on the multinational enterprise and the rationale for foreign direct

³ In terms of the two different approaches to transaction costs identified by Alchian and Woodward (1988) – one concerned with moral hazard problems and monitoring in teamwork, the other with quality and performance of contracts – my concern is firmly with the latter.

investment, and has a substantial pedigree running from Teece (1986) through Kogut and Zander (1992, 1993) to Hallwood (1994). The essence of the matter is the contention that internalization may take place in order to effect knowledge transfer between parties which differ markedly in their capacities for technological absorption, but without the need for opportunistic behaviour by any party to a contractual agreement. This may occur either because some parties are better than others at effecting such transfers of knowledge because of their superior ‘embedded capabilities’ (Kogut and Zander 1993), or because low-knowledge (and therefore high-cost) producers can offer only low rental values for the necessary firm-specific knowledge under a licensing agreement, rendering the establishment of a wholly-owned subsidiary the logical outcome (Hallwood 1994).

Although there has been a debate in the literature on whether this scenario implies a lack of transaction cost issues and/or market failure (Kogut and Zander 1995; Love 1995, 1997; McFetridge 1995; Hallwood 1997), there is some agreement that divergence in technological capability does provide a limited rationale for an opportunism-independent explanation for firm boundaries.

Future Uncertainty

The second opportunism-free approach suggests that contractual difficulties may arise as the result of genuinely different views of an uncertain future. Langlois (1997) argues that even where the parties to a contractual agreement trust each other in virtually every respect, they may have such genuine and irreconcilably different views of what their contractual obligations are under conditions of great uncertainty that joint ownership is the only method of “reconciling divergent visions of the uncertain future” (Langlois 1997:16). This is the strand of the literature echoed by Conner and Prahalad (1996) in their attempt to portray knowledge-based theory as an opportunism-independent theoretical understanding of the firm. Here, even truthful parties may run into contractual difficulties because differences in individuals’ knowledge – which can never be fully known by another party – can lead to radically different judgements about the expected

outcomes arising from a contractual agreement, especially where there is some uncertainty over those outcomes⁴.

While the Langlois approach outlined above puts the emphasis on uncertainty, Madhok (1996) develops a similar theme by using bounded rationality as the key. He argues that bounded rationality appears in two guises: first, an inability to monitor and control behaviour given the likelihood of opportunism, and second, constraints on firm capabilities because they do not possess the necessary knowhow. The latter element is essentially Langlois' 'dynamic transaction costs'⁵: but despite the superficial similarities with Langlois, Madhok suffers from the belief – shared by Kogut and Zander (1993) – that a lack of opportunism implies a lack of market failure (Love, 1995). These differences notwithstanding, there is a clear strand of literature which acknowledges the organizational implications of divergent views of the future which need not depend on contractual parties lying or cheating, but which does acknowledge the implications of self-interested behaviour.

3. A More General Opportunism-independent Theory

The two scenarios outlined above might be regarded as 'special' opportunism-independent theories – they each develop a limited area in which opportunism is not necessary for some aspects of firm organization or structure. The question remains whether it is possible to develop a more general theory which can explain the existence and organization/boundaries of the firm over a wide range of circumstances without recourse to the concept of opportunism as defined earlier.

⁴ Although couched in terms of knowledge differences, Conner and Prahalad's analysis is similar to Rubin's phenomenological approach to the problems of complex contracting. This hinges on bounded rationality arising from the unique 'here-ness' of an individual's personal experience which "...will impel them to violate accepted rules when confronted with a perceived opportunity to benefit from that violation" (Rubin, 1996: 129).

⁵ Langlois describes dynamic transaction (or governance) costs as "the costs that arise in real time in the process of acquiring and coordinating productive knowledge ." (1997: 11).

I argue in this section not only that it is possible to do so, but that two writers have already developed approaches to the firm which, in different but related ways, fulfil this criterion. I do not suggest that either has developed a theory of the firm complete in every respect, but that their work puts us much closer to a general opportunism-independent theory than is contained in the two special cases outlined earlier. One, Ronald Coase, does so by explicitly addressing the problems of market contracting, but parts of his original important message have become somewhat obscured along the way by the edifice of transaction-cost economics erected by his followers. The other, Harold Demsetz, adopts a much more conventionally neoclassical vision of the firm, but nevertheless develops a rationale for both the existence and organization of the firm which is not dependent on opportunism and which puts the costs of knowledge use and acquisition in a central position. Both Demsetz and Coase have been invoked endlessly in the recent literature on the theory of the firm, and especially in the debate on knowledge-based theories and the need for opportunism (Foss, 1996a, 1996b; Conner and Prahalad 1996; Madhok 1996). I contend, however, that the significance of both writers has to some extent been overlooked in this literature, and is in need of reappraisal and extension.

Conner and Prahalad (1996) develop a knowledge-based theory of the firm which they argue is independent of opportunism, and which depends on two key effects: the 'knowledge-substitution' and 'flexibility' effects. It is the first of these with which we are presently concerned. The knowledge-substitution effect "involves the relationship between taking an action and internalizing the wisdom upon which it is based." (Conner and Prahalad 1996:485): in the context of the relationship between manager and subordinate, this involves the manager requiring the subordinate to act according to the former's judgement, rather than the latter's own. This is essentially Demsetz's (1988) concept of 'direction'⁶ i.e. in the context of the employment relation, a manager's knowledge partly substitutes for the employee's knowledge. However, Demsetz's notion of direction has applicability beyond a firm characterized merely as an employment

⁶ Foss (1996b) also makes this point.

contract, and can explain how “the vertical boundaries of a firm are determined by the economies of conservation of expenditures on knowledge.” (1988: 159)

The crux of the argument is to explain how specialized knowledge can be rendered economically useful to those who do not possess it. Learning the knowledge can be rejected as an expensive option which undermines the gains from specialized learning. Buying the knowledge may be possible under some circumstances, but in many cases the economically useful elements of knowledge derive not merely from codified facts but from the tacit theory which underlies those facts, and which cannot easily be transferred by contractual means. Instead, argues Demsetz:

“Those who are to produce on the basis of this knowledge, but not be possessed of it themselves, must have their activities *directed* by those who possess (more of) the knowledge. Direction substitutes for education (that is, for the transfer of the knowledge itself)” (1988: 157-8, original emphasis)

The reason why this process of direction is efficient is spelled out in a later paper (Demsetz 1997), and relies on the existence of economies of scale in the acquisition of specialized knowledge. Direction by those in possession of the knowledge permits the knowledge to be used by guiding the activities of others without the requirement for those others to master the knowledge themselves. This ‘direction’ may be acquired through contractual agreement where it is efficient to do so, or it may be effected indirectly by producing and selling goods which require less information to use than to produce them. In either case, Demsetz is quite clear that this process of determining the (vertical) organization of the firm is independent of opportunism:

“The vertical depth of the firm may be considered from the perspective of the need for conserving on information costs. Other costs matter also, but I wish to focus attention on the consequences of costly knowledge, *and I wish to do so without reference to the information costs inherent in transaction and moral hazard problems*” (1988: 158, emphasis added).

Furthermore, in a thought experiment involving a classic piece of *reductio ad absurdum* Demsetz tackles the important point that it is almost impossible to conceive of the forms of organization which we actually witness in the world around us without appealing to some degree of opportunism. Suppose we have a utopian world where everyone is honourable and there are no unexpected events, and in which there are therefore no gains from the exercise of opportunism. Even in such a world, argues Demsetz, we would still witness vertical integration:

“That this is so follows simply from the fact that any productive activity can be subdivided, and subdivided again and again, so that all activities actually undertaken are, and must always be, vertically integrated to a very considerable degree. There is no unit of product, no unit of production process, and no unit of human activity that is the analog of an indivisible subatomic particle. Opportunism, therefore, cannot explain all or even most of vertical integration” (1997: 21).

Demsetz therefore suggests a theory of the vertical organization of the firm which requires no opportunism or moral hazard, but which nevertheless recognises the costly nature of knowledge acquisition and the economic benefits of knowledge specialization.

There is clearly a connection here with the ‘divergent capabilities’ approach to opportunism-independence outlined above: indeed, one could consider direction itself as a kind of capability, possessed by some economic agents and not others, which helps determine the differential ability to make economic use of specialized knowledge without directly possessing it⁷. But we can take the argument much further. The key issue is whether it is possible to determine an opportunism-independent theory of the *existence* of the firm. For Demsetz (1997: 8-11) the key to the existence of the firm lies in what he calls *specialization*. This has nothing to do with the ‘normal’ meaning of the word, which implies carrying out a narrow range of tasks, but rather the process of producing

⁷ Demsetz’s thesis is sometimes reminiscent of the resource-based/capabilities perspective. Indeed, at times Demsetz is almost Penrosean in his analysis: “Each firm is a bundle of commitments to technology, personnel, and methods, all contained and constrained by an insulating layer of information that is specific to the firm, and this bundle cannot be altered or imitated easily or quickly” (1988: 148).

for others. The firm, therefore, is an entity which is not self-sufficient: *production* (by the firm) is separated from *consumption* (by those outside the firm). Firms exist because producing for others gives rise to efficiency from economies of scale, from the advantages of specialized knowledge (i.e. ‘direction’), and finally because of low transaction costs. The last of these is the inverse of the normal transaction-cost interpretation of firm existence, which depends on high transaction costs. In Demsetz’s world of specialized firms, a reduction in transaction costs causes firms to subdivide and concentrate on an increasingly narrow range of production: the result is more (and more narrowly focused) firms, rather than the substitution of the market for firm activity.

There are three implications of substituting Demsetz’s definition of the firm for the employment-contract relationship envisaged by Conner and Prahalad or by Wernerfelt (1997). First, individuals can be firms. Secondly, firms and markets cannot be regarded as substitutes, since markets do not produce. Rather than having the either/or relationship implied by transaction-cost theory, firms and markets are essentially complements,⁸ because of the productive nature of the former. Thirdly, the one-to-one relationship between ‘the firm’ and the employment contract is broken. For present purposes the first of these implications is trivial, the second is a mixed blessing,⁹ but the third is of considerable importance, for it represents a genuinely opportunism-independent concept of the firm which is not restricted to the employment-contract relationship.¹⁰ This is important, because viewing the firm exclusively in terms of the employment contract not only substantially weakens any claim to have developed an opportunism-independent

⁸ Holmström (1999) also concludes that firms and markets are complements, but does so from a property-rights perspective of the firm as a subeconomy, so that “strong market discipline permits the successful operation of large and powerful firms” (1999: 90).

⁹ A mixed blessing, that is, for transaction cost adherents. Although viewing firms and markets as complements appears to go against a central tenet of transaction-cost theory, it also gets around the tricky issue in transaction-cost analysis of what a zero transaction-cost world would actually look like: if there are no firms, how are things produced in practice? Even Coase has some difficulty in conceiving of the direct contractual relationships which might exist between factors of production and consumers under such circumstances (1993: 65): but a world of no transaction costs poses few problems for Demsetz’s theory of ‘specialized’ firms.

¹⁰ Admittedly, Demsetz has an agenda in developing this definition of the firm. His vision of the specialized firm permits the justification of many of the assumptions underlying neoclassical economics to remain intact, such as the continued focus on productive rather than contracting issues, profit maximizing behaviour, and a reconciliation of the apparently black-box firm of neoclassical theory with something a little more tangible.

theory of firm existence (Foss 1996b), but is a link which Coase himself regrets having made inadvertently, and subsequently rejected (1993: 64-5).

Thus Demsetz develops a form of opportunism-independent theory, one which is not built on market failure arguments but rather on the productive qualities of firms *versus* markets and on the cost-reducing qualities of specialization. The challenge now is to develop an opportunism-independent theory in which the emphasis is more clearly on issues of contracting.

In 1988 the Journal of Law, Economics and Organization published a special issue devoted to a 50th anniversary retrospective of Ronald Coase's 1937 article 'The Nature of the Firm'. One of the benefits of this was that it allowed Coase to reconsider the influence of his 1937 article. As Demsetz (1997: 20) points out, Coase makes it clear in his later writings that he does not believe opportunism offers a special justification for vertical integration: markets involve all kinds of coordination problems which, under some circumstances, may be resolved through management rather than contractual arrangements. Opportunism may exacerbate these problems, but does not lead them into being.

But has Coase anything to tell us about an opportunism-independent theory of firm *existence*? Coase's original purpose was explicitly to explain the existence of firms; by contrast, issues of organization (e.g. vertical integration) were very much secondary and not of great interest to him. The key issue is the cost of coordinating contractual relations between factors of production, which he sees as the plausible alternative to the existence of firms. Coase clearly describes the purpose of 'The Nature of the Firm' as being to show that the basic reason for the existence of firms is the avoidance of the costs of contracting between factors of production (1993: 67). In the absence of firms there would be vast numbers of possible contractual arrangements "...but, absent firms, none would involve the *direction of factors of production*" (1993: 65, emphasis added).

If Coase's idea of 'direction' is the same as that of Demsetz we arrive at a situation in which Demsetz's rationale for the *organization* of firms is the same as Coase's rationale for their *existence*. The original 1937 article certainly couches direction in terms of the employment relation; it is direction in the sense of managers being empowered, within limits, to tell employees what to do. Unlike Demsetz, Coase's version of direction contains no explicit discussion of the problems of tacit knowledge transfer. Rather, the importance of direction for Coase is that it eases the problems of contracting: "The emergence of the firm leads to very much less complicated contractual arrangements" (1993:66). It might be argued that Coase was implicitly allowing for some degree of opportunism in explaining why these arrangements would be complicated, but this appears not to be the case. It is simply the cost of coordinating such a large number of different contracts that poses problems: by 'less complicated' Coase clearly means *fewer* contracts.¹¹ The key issue is that 'for a series of contracts is substituted one' – the opportunity cost of this arrangement (now called transaction costs) is less than for a series of contracts in the open market. There are costs in all this coordination even if there is an absence of opportunism. There may be a less pressing need to draw up and monitor contracts in the absence of opportunism, but there is still the need to coordinate all these arrangements between factors of production, and this involves (opportunity) costs, whether contracts are explicit or not. This is also consistent with Coase's emphasis on search and information costs as a key element in the costs of market contracting¹²: much of the transaction cost literature following Coase has placed an almost exclusive emphasis on the *ex post* problems of bargaining, policing and enforcement, with correspondingly little emphasis on the *ex ante* costs of search and information.

So whether it substitutes for knowledge transfer or eases the problems of contracting, the net effect of direction is the same for both Demsetz and Coase: economic benefits flow from permitting managers to direct the activities of workers. While he expressed this in

¹¹ Despite this emphasis on the complicated nature of market contracting, Coase is a firm believer that markets basically work – see his discussion (1993: 68-72) about asset specificity and the ability of long-term contracts easily to get round apparent problems in the real world, especially on the issue of opportunism being curbed by the market.

terms of the employment contract, Coase subsequently makes it clear (1993: 65) that he erred in appearing to imply that direction should extend only to the employer-employee relationship: it should in principle also extend to the direction of the use of capital (i.e. equipment or money). The concept of direction is therefore vital. For Demsetz it provides an opportunism-free basis for the vertical boundaries and organization of the firm, although not for firm existence. For Coase, direction and the substitution of one contract for many is the essence of the existence of the firm: opportunism may provide an additional incentive for direction to occur but is by no means a necessary condition. This is an opportunism-independent vision of firm existence.

4. Opportunism-independent Contracting and Asset Ownership

In order to develop an opportunism-independent theory of the firm which is nevertheless contractual it must be shown that contracting is meaningful in the absence of opportunism. This section deals with that point, and with the related issue of whether a theory of the firm must look beyond contracts and consider wider issues such as the ownership of assets.

Contractual approaches to the firm fall into two broad types. The first is the nexus of contracts approach which involves an essentially static optimizing problem under conditions of moral hazard and incentive problems, especially where team work is present (Alchian and Demsetz 1972; Jensen and Meckling 1976). The second is more concerned with governance mechanisms and incomplete contracting, and sees firms arising from transaction-cost and institutional factors, or from asset ownership issues (Williamson 1985; Grossman and Hart 1986; Hart and Moore 1990). The Demsetz and Coase visions of the firm outlined above contain elements of both these approaches. Demsetz (1988: 154) is certainly explicit in his view that the firm is properly viewed as a ‘nexus’ of contracts: however, as shown above, in his later work (Demsetz 1997) the emphasis has shifted to explaining the existence of the firm in terms of a ‘specialized’

¹² “The most obvious cost of ‘organizing’ production through the price mechanism is that of discovering what the relevant prices are” (Coase, 1937: 21)

production unit, that is a unit which produces for outside consumption because of the efficiency benefits of specialization, and because of low transaction costs¹³. The Coase approach is quite clearly concerned with the problems of contractual incompleteness and coordination between factors of production, but with the emphasis on the search and information costs involved in finding relevant market prices rather than on the *ex post* contracting problems which characterise Williamson's later developments.

But why do we need contracts at all if there is no threat of opportunism? The issue here revolves around what we think contracts are *for*, and thus what we actually mean by a 'contractual' approach to the firm. Some writers link the term 'contractual' by definition to opportunism/moral hazard: "the very notion of contract is hard to rationalize in lieu of the notion of opportunism" (Foss 1996b: 522). If opportunism is a necessary condition for contracting, and contracting is the essence of the existence/boundaries of the firm, then the search for an opportunism-independent theory is indeed doomed to failure. But even if contracts are simply defined as a form of promise (i.e. commitment today against performance tomorrow), their existence need not depend on opportunism. Possible functions of a contract are not only to guard against opportunism, but to give guidance about what to do in the event of an uncertain future (Langlois, 1997) or to eliminate or attenuate misunderstandings of fact (Alchian and Woodward's 'honest disagreements'). In the absence of such guiding mechanisms behaviour may develop which seems perfectly reasonable to one party but may be construed as being quite unreasonable by another, especially under conditions of uncertainty. Such behaviour may not be 'opportunistic' in the Williamsonian sense, because it need not be characterised by the presence of guile or deceitfulness.

And contracts may have functions which go beyond even these roles. In an analysis of competence development and transfer in the UK offshore oil and gas industry, Finch

¹³ This is not to say that Demsetz has abandoned his belief in the similarity of the authority relationships between employer and employee and supplier and customer so famously presented in Alchian and Demsetz (1972). "[I]f the contractual arrangements between a firm and its suppliers share characteristics with the contractual arrangements between a firm and its workers, and if these characteristics respond to the same types of forces, why should we distinguish between them?" (1997: 16). But here his concern is with the internal organization of the firm rather than its existence.

(2002) examines the nature of the relationships between the operating companies which carry out exploration and development activity and the numerous contracting companies which carry out a range of specialist (mainly engineering) tasks for the operating companies. While Finch's objective is to determine the range of relationships between these companies and how this has affected the transfer and development of competences in the sector, he makes some relevant observations on the nature of the contractual relationships between them. These involve the full gamut of supply-chain management issues such as alliances and partnerships as well as more traditional forms of contracting. While contractual arrangements do involve designing appropriate incentives and performance indicators, they frequently go far beyond this, providing the means to permit 'cultural alignment' between companies with different business objectives, and even permitting the optimal level of integration for devolving responsibility and allowing appropriate competence transfer from operating company to contractor. Crucially, Finch finds that contracts have an important role to play in this sector not merely to protect against wrongdoing by contractual parties, but to provide the necessary encouragement for new technical solutions to engineering problems, and permitting the framework within which (personnel-based) competence transfer may occur. He points out that this is a highly knowledgeable community in which (guileful) opportunism is difficult to enact: but there are still contracts.

This suggests that seeing contracts entirely as defence mechanisms where there exists a threat of opportunism may be an unduly narrow vision: the question is not only why we should have contracts where there is no opportunism in theory, but why it is that we continue to see contracts being used even where opportunism is unlikely to be a serious economic threat in reality.

I began the discussion of the opportunism-independent view of the firm by suggesting that the firm in theory was best seen in terms of a multi-person entity, but which is more than merely an employment contract. It has been argued that explaining the firm in terms of an employment contract is insufficient, and that a complete theory also has to be able to offer an explanation of the ownership of (physical) assets. The true challenge, then, is

to find a rationale for the ownership of assets by firms without the need to appeal to opportunism/moral hazard (Foss 1996b). This in turn poses the question of whether the theories of the firm developed by either Coase or Demsetz deal satisfactorily with this issue.

The importance of asset ownership arises from the incomplete contracting or property rights approach (Grossman and Hart 1986; Hart 1989; Hart and Moore 1990). The boundaries of the firm are here determined by the efficient allocation of ‘residual rights’, the rights to make decisions about production which are not made explicit in any contract. Where the assets brought by two parties to a contractual agreement are highly specific, and where opportunism is therefore a credible threat, efficiency may be improved by giving one party ownership of both sets of assets, thus vesting the residual rights with that party. This in turn, it is argued, not only defines the firm clearly as a bundle of assets under common ownership, but also tells us something about the nature of the relationship between employers and employees, because employers have the residual rights with respect to the physical assets which determine the employees’ productivity.

Although, as indicated earlier, Coase spends some time indicating that his concept of the firm extends beyond employment contracting and includes “the contracts that enable the organizers of the firm to direct the use of capital (equipment or money) by acquiring, leasing, or borrowing it” (1993: 65), he is generally unspecific about the nature of such contracts. Again the issue relates to the gains to be made from substituting one contract for many, rather than specifically overcoming opportunism or moral hazard issues by replacing market contracting with firm organization.¹⁴ Demsetz too spends little time on this issue, but when he does address asset ownership he clearly does not regard it as being central to the essence of the firm. In his discussion of the knowledge and monitoring benefits derived from the firm acting as a ‘central contracting agent’ (1997: 33-34), Demsetz rejects Oliver Hart’s (1989) argument that asset specificity and residual control rights are important in determining the identity of the central contracting agent. For

¹⁴ This is not to say that Coase ignores entirely the issue of asset ownership. He is perfectly clear in discounting specific assets as a likely explanation for vertical integration (1993: 68-71).

Demsetz, asset ownership follows the benefits of specialization in knowledge, rather than *vice versa*:

“Ownership of an asset is not predetermined. It should gravitate to the party who possess (sic) the specialized knowledge necessary to do a good job of setting the general goal of the cooperative effort and monitoring the degree to which various owners of inputs accomplish their tasks” (1997: 34, fn 15).

However, that neither Coase nor Demsetz pays particular attention to issues of asset ownership need not imply that an opportunism-independent vision of the firm is deficient in a crucial area. As two recent and influential contributions have argued, the Hart/Moore property rights approach fails to offer a complete explanation of the boundaries of the firm, and the importance of asset ownership is overstated in terms of a theory of the firm. In the first of these, Wernerfelt (1997) tackles the issue of why employees submit to the authority of their employers. This was initially raised by Alchian and Demsetz (1972) in their suggestion that there was no difference between authority in employment and (contractual) ‘authority’ in a market relationship, and is resolved in the property rights approach by the firm’s ownership of the residual rights to productive assets. However, Wernerfelt points out that authority exists even in circumstances where there is little room for the kind of incentive conflicts which give rise to the firm in the property rights analysis, such as members of volunteer organizations or the crew of a racing yacht. If incentive conflicts can be separated from authority *per se*, this suggests that submission to authority can be an efficient communication pattern where there are no incentive conflicts, and even where such conflicts exist as long as the parties can settle possible conflicts up front. Wernerfelt then goes on to develop a model which explains when each of three possible types of contracting for human asset services will be most efficient,¹⁵ based on the communication costs of adjustment involved in each. But for present purposes the importance of Wernerfelt’s analysis lies in showing that there is no necessary link between asset ownership and residual rights on one hand, and authority relations (or ‘direction’) on the other.

¹⁵ The three types of contract are the hierarchy form (i.e. an employment relationship), the negotiation-as-needed form, and the price list form.

In the second contribution, Holmström (1999) argues that that the property rights approach does not explain why firms own assets, indeed that it says very little about firms at all: rather it is a theory of asset ownership by individuals. The Hart-Moore model, Holmström argues, is subject to precisely the same critique levelled at Alchian and Demsetz (1972), “that organizational affiliations did not matter for transactions” (1999: 87). This is important, because individual ownership of assets does not provide a theory of the firm *unless* individuals are firms¹⁶: so the Hart-Moore approach is not a theory of firm boundaries, but an entrepreneurial theory of the firm. “The problem is that there are really no firms in these models, just representative entrepreneurs” (Holmström, 1999: 100). This in turn matters because concentrating on issues of asset configuration as an end in itself is less important than the activity configurations which these imply: the important issue is what firms *do* rather than what they *own*.

Holmström then goes on to suggest alternative reasons as to why firms rather than workers generally own nonhuman productive assets, and does so in a way which echoes the wider interpretation of the functions of contracts outlined earlier in Finch’s work. By having assets under a single authority, firms can assign workers in a manner which is much more varied than under separate ownership. The firm can promote, dismiss or move workers in any way it deems fit, but only within the set of assets which it owns: it is a form of internal capital market. Crucially, “[b]y focusing on holdups alone, the property rights approach overlooks the great variety of instruments that can be used to influence employee incentives” (Holmström, 1999: 89), and these incentives go far beyond the simple ones of bonus payments and so on, but include much more subtle arrangements such as the control of information channels, the assignment of tasks, and even espousing a particular corporate culture. Thus the leverage which the firm has over its human assets via the holding of contracting rights over nonhuman assets arises not merely from dealing with the holdup power of employees, but from the range of instruments which the firm has at its disposal to deal with these issues.

¹⁶ As noted earlier, Demsetz is unlikely to regard this as a problem because he explicitly permits the possibility of individuals as firms.

This discussion of the wider functions of contracts and of the property rights approach to firm boundaries therefore leads to two conclusions. First, contracting is indeed meaningful within an opportunism-free environment, but the functions performed by contracts must be viewed as going beyond those of incentive alignment designed to protect against wrongdoing. Secondly, the need for an opportunism-independent theory to include an explicit explanation for the firm's ownership of non-human productive assets is easily overstated: not only does the possession of residual rights to such assets fail to explain the existence of authority in many circumstances (Wernerfelt 1997), but as Holmström indicates, the property rights model fails as a theory of the *firm*.

5. Towards an Opportunism-independent Theory?

Although their respective rationales for the existence of firms differ somewhat, there are clear similarities between Demsetz's and Coase's conceptions of the firm. First, both place substantial important emphasis on 'direction', the process by which one set of economic agents is permitted to exercise control over the activities of human and (in the case of Coase) non-human factors of production. Secondly, neither sees opportunism as necessary for either firm organization or firm existence: neither Demsetz nor Coase appears to have any problems in conceiving of a contractual world in the absence of opportunism, a world in which firms may exist either, as in the case of Demsetz, because of the benefits of 'specialization' (that is, separating production and consumption), or, as in the case of Coase, because there are savings to be made under some circumstances by replacing many contracts with one through the process of 'direction'. Thirdly, both Demsetz and Coase have a contractual view of the firm, Demsetz from a 'nexus of contracts' perspective, Coase from the perspective of the problems inherent in market contracting. Finally, neither author puts much store on ownership of assets as a means of defining the boundaries of the firm.

This takes us far beyond the 'divergent capabilities' and 'uncertain future' approaches discussed earlier. It is perfectly possible to conceive of the firm as a contractual entity,

but one which does not necessarily hinge on moral hazard/opportunism issues. For example, 'direction' is not about rent generation and strategic behaviour under all circumstances: the work of Demsetz and Coase highlighted above indicates that direction can have an important influence on the choice of governance structures in the absence of strategic behaviour on the part of economic agents. Crucially, an opportunism-independent theory of the firm is not dependent on a knowledge or resource-based perspective (Conner and Prahalad 1996; Foss 1996a,b): such a conception of the firm is entirely consistent with the neoclassical and transaction-cost perspectives.¹⁷

But while we have moved towards an opportunism-independent yet contractual theory of the firm, the task is far from complete. For example, Coase is by no means always precise in explaining the particular circumstances under which direction, and replacing many contracts with one, is likely to be efficient. A clearer method of determining these circumstances is required, with one possible starting point being the adjustment-cost approach of Wernerfelt (1997). Wernerfelt shows formally that the choice between the general employment relationship in which detailed tasks are not fully specified in advance and other forms of contracts (such as 'negotiation-as-needed' or set 'price lists' for services) depends crucially on the extent of adjustments needed in the contracted-for service, and the communication costs (rather than incentive conflicts) resulting from this. Although limited by viewing the firm exclusively as an employment relationship, theoretical work of this type is helpful in moving away from the view of contracts merely as incentive-aligning devices. This relates to the work of both Finch and of Holmström reviewed earlier, which concentrates on the role of contracting beyond that of incentive alignment merely to prevent malfeasance. Holmström's approach is certainly about incentives, primarily in the employment relationship, but stresses that merely paying employees based on their holdup power greatly underestimates the variety of instruments which firms possess to influence workers. Finch stresses the nature of contracting relationships with those outside the firm, and shows that contracting can play a role in competence transfer and development where the exercise of opportunism is not a serious

¹⁷ As even Demsetz is prepared to admit, the information/knowledge costs on which he puts so much emphasis are an important element in transaction costs (1988: 141), which is why he stresses information costs in discussing the benefits of specialization and direction.

threat. This suggests an avenue of research exploring in much greater detail the possible roles of contracting beyond those normally considered in an incomplete contracting setting, possibly complementing the considerable insights which we now have into situations in which we do *not* find contracts when we might expect to do so, such as R&D cooperation (Buckley and Chapman 1997, 1998).

In addition, if we are to consider fully a world in which firms contract in the absence of Williamsonian opportunism, then it pays to be better informed of the conditions under which opportunism is likely to arise. This relates to the point raised earlier that those interested in transaction cost analysis have placed much more emphasis on bargaining and decision costs and policing and enforcement costs, rather than the *ex ante* search and information costs stressed by Coase (1937). Stephen (1996) points out that bargaining and policing costs may themselves be a function of which party is selected as a potential contracting partner, so that if it is possible to engage in search to discriminate *ex ante* between potential contractors with a high likelihood of opportunism and those with a low likelihood, there may be a trade-off between search costs and enforcement costs. Stephen shows that *ex ante* screening can be an effective substitute for *ex post* contractual enforcement in the case of staff selection and evaluation in the legal profession (Stephen and Gillanders 1993), and suggests that where it is possible for potential contractors to signal credibly the appropriate characteristics, such screening devices could be used elsewhere, such as in prequalification requirements in the oil industry. This suggests a two-stage empirical research agenda; first, determine the extent to which *ex ante* screening is a credible alternative to *ex post* policing and enforcement in certain settings as a means of avoiding opportunistic behaviour, and secondly, where the threat of opportunism is very low or absent, examine the rationale for the continued existence of contracts under such circumstances.

In summary, my argument is that a theory of the firm which emphasizes contractual relations need not depend on Williamsonian opportunism. I do not, however, suggest that an opportunism-independent theory of the firm *must* be contractual. There are certainly dangers in concerning ourselves too much with a vision of the firm which depends

wholly on contracts and the process of exchange, dangers well recognised by both Coase and Demsetz. In quite different ways, both are concerned with the ability of firms to act as more than merely exchange mediums or contractual conduits¹⁸. For Demsetz this is clear from his insistence that the essence of the firm is its “productive arrangements” (1997: 4) and his view that the difference between markets and firms is that markets do not *produce* anything. Coase expresses similar concerns about the way in which transaction-cost theory has developed. It has:

“..encouraged an undue emphasis on the role of a firm as a purchaser of the services of factors of production and on the choice of contractual arrangements which it makes with them. As a consequence...economists have tended to neglect the main activity of a firm, *running a business*” (1993: 65, emphasis added).

Coase’s warning is highly apposite when searching for any construct of the firm which is actually useful either for economists or strategy writers. For example, possibly the major virtue of the resource-based approach is that it reminds us that firms are more than merely the sum of the contracts which determine their boundaries, but are entities which can create value - not merely reduce (transaction) costs - in a way that would not happen in the market (Conner, 1991). Increased emphasis on ‘running a business’ is therefore to be welcomed, but this does not mean that we must abandon conceiving of the firm as a contractual entity.

¹⁸ Coase and Demsetz are not alone in these concerns. Kay (2000) also argues that there has been an undue emphasis on contractual relations in the theory of the firm, to the neglect of *decisions*. “Firms can make decisions, markets can only stimulate and inform decisions...A firm can allocate resources without a market but markets cannot allocate resources without firms.” (p. 696)

References

Alchian, A. A. and Woodward, S. 1988. The firm is dead; long live the firm. A review of Oliver E. Williamson's *The Economic Institutions of Capitalism*, *Journal of Economic Literature*, vol. 26, 65-79.

Buckley, P. J. and Chapman, M. 1997. The perception and measurement of transaction costs, *Cambridge Journal of Economics*, vol. 21, 127-45.

Buckley, P. J. and Chapman M. 1998. The management of cooperative strategies in R&D and innovation programmes, *International Journal of the Economics of Business*, vol. 5, 369-81.

Coase, R. H. 1937. The nature of the firm, *Economica*, vol. 4, 386-405.

Coase, R. H. 1993. The nature of the firm: influence, in Williamson, O. E. and Winter, S. G. (eds), *The Nature of the Firm: Origins, Evolution and Development*, Oxford, Oxford University Press.

Conner, K. R. 1991. A historical comparison of resource-based theory and five schools of thought within industrial organization economics: do we have a new theory of the firm?, *Journal of Management*, vol. 17, 121-154.

Conner, K. R. and Prahalad, C. K. 1996. A resource-based theory of the firm: knowledge versus opportunism, *Organization Science*, vol. 7, 477-501.

Demsetz, H. 1988. The theory of the firm revisited, *Journal of Law, Economics and Organization*, vol. 4, 141-61.

Demsetz, H. 1997. *The Economics of the Business Firm: Seven Critical Commentaries*, Cambridge, Cambridge University Press.

Finch, J. H. 2002. Transferring exploration and production activities within the UK's upstream oil and gas industry: a capabilities perspective, *Journal of Evolutionary Economics*, vol. 12, 55-81.

Foss, N. J. 1996a. Knowledge-based approaches to the theory of the firm: some critical comments, *Organization Science*, vol. 7, 470-76.

Foss, N. J. 1996b. More critical comments on knowledge-based theories of the firm, *Organization Science*, vol. 7, 519-23.

Fourie, F. C. v N. 1993. In the beginning there were markets? in Pitelis, C. (ed) *Transaction Costs, Markets and Hierarchies*, Oxford, Blackwell.

Goldberg, V. P. 1984. A relational exchange perspective on the employment relationship, in Stephen, F. H. (ed) *Firms, Organizations and Labour*, Macmillan, London.

Grossman, S. and Hart, O. 1986. The costs and benefits of ownership: a theory of lateral and vertical integration, *Journal of Political Economy*, vol. 94, 691-719.

Hallwood, C. P. 1994. An observation on the transaction cost theory of the (multinational) firm, *Journal of Institutional and Theoretical Economics*, vol. 150, 351-61.

Hallwood, C. P. 1997. The transaction cost theory of the (multinational) firm: reply, *Journal of Institutional and Theoretical Economics*, vol. 153, 682-87.

Hart, O. 1989. An economist's perspective on the theory of the firm, *Columbia Law Review*, vol. 89, 1757-74.

- Hart, O. and Moore, J. 1990. Property rights and the nature of the firm, *Journal of Political Economy*, vol. 98, 1119-58.
- Hodgson, G. M. 1993. Transaction costs and the evolution of the firm, in Pitelis, C. (ed) *Transaction Costs, Markets and Hierarchies*, Oxford, Blackwell.
- Holmström, B. 1999. The firm as a subeconomy, *Journal of Law, Economics and Organization*, vol. 15, 74-102.
- Jensen, M. and Meckling, W. 1976. Theory of the firm: managerial behavior, agency costs and ownership, *Journal of Financial Economics*, vol. 3, 305-60.
- Kay, N. 2000. Searching for the firm: the role of decision in the economics of organization, *Industrial and Corporate Change*, vol. 9, 683-707.
- Kogut, B. and Zander, U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology, *Organization Science*, vol. 3, 383-97.
- Kogut, B. and Zander, U. 1993. Knowledge of the firm and the evolutionary theory of the multinational corporation, *Journal of International Business Studies*, vol. 24, 625-45.
- Kogut, B. and Zander, U. 1995. Knowledge, market failure and the multinational enterprise: a reply, *Journal of International Business Studies*, vol. 26, 417-26.
- Langlois, R. N. 1997. Transaction costs, production costs, and the passage of time, in Medema, S. G. (ed) *Coasean Economics: Law and Economics and the New Institutional Economics*, Dordrecht, Kluwer.
- Love, J. H. 1995. Knowledge, market failure and the multinational enterprise: a theoretical note, *Journal of International Business Studies*, vol. 26, 399-407.

Love, J. H. 1997. The transaction cost theory of the (multinational) firm: note, *Journal of Institutional and Theoretical Economics*, vol. 153, 674-81.

McFetridge, D. G. 1995. Knowledge, market failure and the multinational enterprise: a comment, *Journal of International Business Studies*, vol. 26, 409-15.

Madhok, A. 1996. The organization of economic activity: transaction costs, firm capabilities and the nature of governance, *Organization Science*, vol. 7, 577-90.

Masten, S. E., Meehan, J. W. and Snyder, E. A. 1991. The costs of organization, *Journal of Law, Economics and Organization*, vol. 7, 1-25.

Rubin, E. L. 1996. The phenomenology of contract: complex contracting in the entertainment industry, *Journal of Institutional and Theoretical Economics*, vol. 152, 123-39.

Stephen, F. H. 1996. Lawyers, transaction costs and opportunism, *Journal of Institutional and Theoretical Economics*, vol. 152, 146-53.

Stephen, F. H. and Gillanders, D. D. 1993. Ex post monitoring versus ex ante screening in the new institutional economics, *Journal of Institutional and Theoretical Economics*, vol. 149, 725-30.

Stephen, F. H. and Love, J. H. 2000. Hold-up costs, economies of scale and the make-or-buy decision, Research Paper RP0020, Aston Business School, Birmingham.

Teece, D. 1986. Transaction costs economics and the multinational enterprise: an assessment, *Journal of Economic Behavior and Organization*, vol. 7, 21-45.

Wernerfelt, B. 1997. On the nature and scope of the firm: an adjustment-cost theory, *Journal of Business*, vol. 70, 489-514.

Williamson, O. E. 1985. *The Economic Institutions of Capitalism*, New York, Free Press.

Williamson, O. E. 1993. Opportunism and its critics, *Managerial and Decision Economics*, vol. 14, 97-107.