ABSTRACT
The contractual theory of the firm, developed by Ronald Coase (1937), Oliver Williamson (1975, 1985a) and others, posits the firm as an efficiency-enhancing nexus of contracts where several input owners make bilateral contracts with a central agent – an individual or group entrepreneur. The focus is on contracts within the sphere of production, to the neglect of the legal personality of the firm. This makes it difficult to come to terms with the facts that firms typically own the assets used in production, and that firms, rather than their employees or owners, make contracts with customers or suppliers. The reinstatement of the concept of legal personality in the theory of the firm is necessary to account for these real-world facts. The efficiency-enhancing central contractual agent is the separate legal person in which ownership rights over assets used in production are vested, in whose name contracts are made, and thanks to which firms have access to courts. It is this singular legal personality that makes the firm an integrated whole, with the unitary capacity to enter into contracts with others, and to sue and be sued. The paper calls for a view of the firm that recognizes in law’s provision of legal-entity status a fundamental institutional support for the firm.
This essay considers the nature and definition of the firm. After an outburst of innovative research on the theory of the firm in the 1970s, there has been relatively little theoretical development in this area since 2000. A problem is that economists cannot agree what a firm is. Different arguments for the existence and productivity of the firm often cannot readily be tested conjointly, partly because of a lack of clear definitions and concepts. This is a chronic problem with empirical work on the theory of the firm (Carter and Hodgson 2006).

This failure to reach a consensus on definitions has multiple causes. Among them is a failure to acknowledge that property, contracts, firms and corporations are all historically specific and relatively recent phenomena. The story of how economics transformed itself into an ahistorical discipline, and elevated notions such as property, contract and exchange into eternal verities, while simultaneously eroding them of much meaning, is told elsewhere (Hodgson 2001). If firms and contracts are regarded as universal to all human existence then we are unable to treat them as outcomes of relatively recent, state-based, legal systems.

Both Ronald Coase (1937, 1988) and Oliver Williamson (1975, 1985a) made little of the legal personality of the firm. Williamson treated law as if it were akin to custom, with arrangements between parties as ‘private ordering,’ which could in principle emerge without the involvement of a state legal system. There has been insufficient acknowledgement of the role of the state in bringing corporations into existence, and of how the development of company law stimulated entrepreneurial organizations that drove much of the explosive growth of capitalism in the last two hundred years.

It is argued here that firms in general (including the particular legal form of the corporation) have to be treated as legal entities, where law itself is irreducible to custom or private ordering. We use the term firm to apply to organizations that are functioning legal units set up to produce goods or services for sale. It can be stretched to cover individual producers as well. The Concise Oxford English Dictionary defines the firm as ‘partners carrying on business; group of persons working together.’ A partnership is one type of firm, where individuals enter into agreements to produce jointly. A corporation is yet another kind of firm, structured

---


2 Among the foremost twenty results from a Google Scholar search using ‘theory of the firm’ (performed 22 December 2014) only two items originated from 2000 or later. None of these two was from a leading journal of economics.
as designated under corporate law. All corporations are firms, but not all firms are corporations.3

The word firm derives from the Latin adjective firmus, meaning strong, powerful, durable and lasting. As a noun, the word went on to acquire the significant meaning of a (legally binding) signature and with this important connotation it survives today in some form in several Romance languages. It is reasonable to use the broad term firm to include a both partnership and a corporation, as well as a single person employing others in production or trade. But, of course, these different types of firms all have important differences of structure, rationale and possible behaviour. Nevertheless, they can all be placed within a single taxonomic class. Within these inclusive terms, a more precise definition of the firm will be attempted later.

We are less concerned with theories that explain specific firm-related phenomena, such as vertical integration, firm boundaries, or firm structure. Instead this selective survey is concerned with what leading authors mean by firm and whether their account of this entity is coherent or robust. Section 1 shows how the conception of a firm in transaction cost economics slid from a firm-market dichotomy to a firm-market continuum. Section 2 shows that the notion of a firm-market hybrid is untenable once we adopt the legal conception of the firm, which is missing in transaction cost economics. Section 3 considers some other major contributions to the theory of the firm since 1970. Section 4 concludes the essay.

1. The firm in transaction cost economics

In his famous article Coase (1937) treated the firm and the market as two alternative ways of organizing productive activity. The firm was defined in terms of its ‘supersession of the price mechanism’ (p. 389). For Coase (p. 391), owners of factors of production do ‘not have to make a series of contracts’ with other factor owners in the production process: ‘For this series of contracts is substituted one.’ Each factor owner makes a contract with the ‘entrepreneur,’ which was defined as ‘the person or persons who … take the place of the price mechanism in the direction of resources’ (p. 388 n.). ‘A firm … consists of the series of relationships which comes into existence when the direction of resources is dependent on the entrepreneur’ (p. 393).

Coase’s account depends on several legal concepts, including owner, sale and contract. But a major flaw is that the legal constitution and role of the (individual or collective) ‘entrepreneur’ – who owns resources, makes contracts, and directs production – is incompletely specified. Unless this problem is fixed, his distinction between the firm and the market disintegrates.

Coase (1937) regarded the entrepreneur as one or more people. Assume first that it is one person, who enters into legal contracts with employees and suppliers of materials, and that production takes place. Clearly this firm is a legal entity, constituted by a real person with the

---

3 In contrast, in an otherwise excellent article, Robé (2011, p. 3) argued that they are ‘totally different concepts: a corporation is a legal instrument, with a separate legal personality, which is used to legally structure the firm; a firm is an organized economic activity, corporations being used to legally structure most firms of some significance.’ But the idea that the firm is an ‘activity’ is odd: most writers treat the firm as a productive organization. Robé (2011, p. 10) stated – with little review of meanings and relevant arguments in the literature – that ‘it is the corporation which has legal personality and not the firm.’ He did not acknowledge that the concept of the firm is widely used in the different sense of a productive organization, of which corporations are but one example.
intention of producing goods and services and with the legal capacity to make contracts. But Coase did not clearly acknowledge that the entrepreneur owns the product, and has the right to the revenue from the goods or services that are produced. Instead of entrepreneurial ownership rights and potential liabilities, he concentrated on the administrative functions of the entrepreneur during the production process, including his claim that the employment contract involves flexible but limited authority over employees (p. 391).

Because Coase concentrated on the administrative functions of the entrepreneur within production, he overlooked another important issue. Who would be sued if the output of the firm proved defective or dangerous? Would it be the entrepreneur, or the individual worker responsible for the defect? With a legally-constituted firm it would be impossible for an outsider to sue the culpable worker directly – the firm would be sued. The legal formation of a firm establishes it as the locus of legal liability in trading with others. But because Coase is insufficiently clear on the legal owner and seller of the output, he misses this point entirely.

These problems are compounded when the entrepreneur is more than one person. For instance, what are the ties or incentives that keep the entrepreneurs together as a team? Coase (1937) was again silent on this point. But there must be some legally enforceable arrangement that keeps them together, and allows them as a body to make the contracts with the owners of factors of production. Employees and suppliers would not make separate contracts with each of the entrepreneurs as individuals, so with whom are they contracting? Likewise, if the firm sells a defective product, then which entrepreneur gets sued? Or are they sued as a body? In which case, how would the liability be shared between them?

If the entrepreneurs were legal partners, then they could be bound together by a legal partnership agreement or articles of partnership. This agreement would specify management responsibilities, shares of profits or losses, and mechanisms to resolve disputes between partners. Unlike a corporation, partners are jointly and severally liable for the full value of any partnership debts. Suing a partnership means suing the partners. The glue holding the firm together, and making it a singular unit, is the partnership agreement of joint responsibility. Contrary to Coase, the partnership firm is constituted not by entrepreneurial administration of a production process, but by the legal presumption that those partners, who join forces to pursue entrepreneurial activities, also share the relevant responsibilities.

In a corporation, entrepreneurs could refer to shareholders or managers. Legal incorporation means that the state recognizes the firm as a singular legal person, with rights and duties. The corporation does not consist solely of its entrepreneurs. Neither managers nor shareholders own the corporation. The corporation itself is an owning agent. Shareholders own shares in the corporation: they do not own the corporation itself (Marris 1964, p. 12, Gower 1979, Ireland 1999, Blair and Stout 1999, Robé 2011). The corporation, as a legal person, hires the workers, buys machines and raw materials, and sells the output. It can be sued if it sells defective products. It can sue others for breaches of contract. The glue binding the corporation together is the power of corporate law, the adoption of its principles by the shareholders, and the legal agreement between them. Contrary to Coase, the corporation is not constituted by entrepreneurial administration of a production process, but by the establishment of the singular legal person that can combine the agency of multiple individual entrepreneurs.

Note that there are many corporations with only a single shareholder, including corporations owned by another corporation.
Eventually, his neglect of the legal personality of the firm meant the loss of a strict firm-market dichotomy. Even in his classic article, when Coase (1937, p. 388) pointed out that in a department store there may be ‘competitive bidding for space’ or that in the Lancashire cotton industry ‘a weaver can rent power and shop-room and can obtain looms and yarn on credit,’ these might be conceived as markets within firms, leading to ‘hybrid forms.’ But, once we conceive of the firm as a legal entity, then internal markets and hybrids disappear. A franchisee in a department store is legally separate from the franchising department store itself, and the Lancashire weavers to which Coase referred were self-employed producers, even if they worked inside a mill owned by another firm. Because Coase did not conceive of the firm as a legal entity, the rot set in. Then Benjamin Klein (1983, p. 373) pushed against Coase’s fragile framework, and previous distinctions collapsed:

Coase mistakenly made a sharp distinction between intrafirm and interfirm transactions, claiming that while the latter represented market contracts the former represented planned direction. Economists now recognize that such a sharp distinction does not exist and it is useful to consider also transactions occurring within the firm as representing market (contractual) relationships. The question what is the essential characteristic of the firm now appears to be unimportant.

Here, Klein made an unwarranted appeal to authority and advanced a non sequitur. Despite the mention of unnamed ‘economists,’ researchers will search with difficulty in the academic literature for any forceful argument against the ‘sharp distinction’ between the firm and the market. There is no known scholarly explanation why it would be ‘useful’ to consider transactions within the firm as ‘market’ relationships. No convincing argument has been found for abandoning the question of ‘the essential characteristic of the firm.’ Even if the ‘sharp distinction between intrafirm and interfirm transactions’ were untenable, this does not mean that we can abandon the question what is the ‘essential characteristic of the firm,’ because distinctions would still have to be made between degrees of ‘firm-ness’ and ‘market-ness.’ If we believe that the firm has no clear identity, then it becomes a conceptual sponge ready to soak up anything put in contact with it.

Then Coase (1988, pp. 27-28) revised his position: ‘I have come across numerous examples of markets found within firms, but one which amused me was the discovery of a kind of market operating in the heart of a nationalized industry in England, the electricity supply industry.’ He quoted from a 1961 lecture given by an official of the Central Electricity Generating Board (CEGB): ‘the National Control Room becomes in effect an auction room, with a National Control Engineer asking the Regional Centres to quote the price at which they could supply a certain number of kilowatts at specified periods during the following day. … Wherever possible he accepts the lowest bid.’ Coase continued: ‘An analogous situation may, of course, be found within a privately owned firm in which separate departments or divisions may supply one another as a result … of what are essentially market transactions between them.’ The firm-market dichotomy had disappeared.

But, once we apply clear legal criteria concerning contracts and firm ownership, Coase’s argument disintegrates. The UK CEGB was itself a singular legal person. It was a state-owned company, lasting from 1957 until its privatization and division into separate companies in the 1990s. The regional centres were not legally separate firms but internal divisions of the CEGB. An outside complainant would sue the CEGB, not the regional centres, just as the CEGB would sue any supplier of materials to a regional centre that did not fulfil a contract. The regional centres neither owned nor sold electricity to the CEGB. Any semblance of contract between the regional centres and the CEGB would not have been recognized in law.
as such. It was not an exchange of property rights because ownership of the electricity remained in the hands of the CEGB throughout. Instead, the bidding and ‘selling’ of electricity was an internal management mechanism to reduce costs and encourage increases in productivity.

There are often internal negotiations and transfers of resources between divisions of modern firms. These divisions may have their own accounts and profit targets. Most sizeable firms use price indicators for internal accounting. But are there internal markets within firms? Again a key test is whether these divisions have separate legal status, and are recognized as legal persons. Internal transfers within the firm do not involve the exchange of legal property rights. The objects of ‘exchange’ remain the property of the firm. These ‘exchanges’ are not legally enforceable contracts of trade: they are internal transfers. If a division of the firm is delegated the power to enter into contracts with outside bodies, then the firm as a whole is legally the party to the contract. The division is merely exercising delegated powers: it acts in the name of the corporation, and the corporation as a whole is legally responsible for any liabilities under the agreed contract. Because the firm is a singular legal entity, it cannot make contracts within itself, just as our legs cannot make a legal contract with our mind to walk or run when instructed.

Williamson also moved from the dichotomy to the continuum. Although his seminal work *Markets and Hierarchies* (1975) suggested a dichotomy, it was soon to disappear. Like Coase, Williamson (1981, p. 1538) chose to concentrate on ‘the internal organization of the corporation’ and to downplay its legal personality. Williamson (1985a, p. 318) noted that ‘the centrality of management … distinguishes it from all other constituencies.’ Williamson (1985b, p. 199) further explained: ‘whereas each constituent part of the enterprise strikes a bilateral deal with the firm … management has knowledge of and is implicated in all of the contracts.’ But this formulation runs into the very same problems that we have observed with Coase. What binds management, the officers and directors of a corporation, together? Being ‘implicated in all of the contracts’ is not the same as identifying the legally responsible entity.

More preoccupied than Coase with private ordering, Williamson (1985b, p. 184) wrote: ‘Since the efficacy of court ordering is problematic, contract execution falls heavily on the institutions of private ordering … This is the world with which transaction cost economics is concerned.’ This suggested, without much supporting argument, that private ordering is generally much more effective than court ordering. For this questionable reason, Williamson proposed that transaction cost economics should overlook courts and statutory law. This severe confinement of the ‘world with which transaction cost economics is concerned’ is a serious impairment for this research program, and it is fatal for attempts to establish a clear identity for the firm.

Like Coase, Williamson treated the firm as a group of individuals, such as partners or shareholders. But this is insufficient to integrate the firm a cohesive entity, or to define its boundaries. Consequently, distinctions between the firm (or ‘hierarchy’) and the market faded away. Williamson (1985a, p. 83) became ‘persuaded that transactions in the middle range are much more common.’ For Williamson (1991, p. 271), hierarchies (or firms) became ‘a continuation of market relations by other means.’ Instead of the firm-market dichotomy, he adopted a firm-market continuum.

Richard Langlois (1995, p. 72) observed that ‘much of transaction-cost economics has reached the conclusion that the distinction between firm and market is little more than semantics.’ In ‘transaction cost economics’ the theory of the firm has become a theory of different types of individual contractual arrangements and their consequences. Ironically, in
the age of the large corporation, the firm as such has virtually disappeared from standard economic analysis. Typically, the firm is defined not as an entity, but as a point on a continuum of possible governance and contracting structures. As Williamson (2007, p. 376) put it:

What defines a firm at the end of the continuum? I take the defining characteristics of governance structures to be incentive intensity, administrative control, and the contract law regime. Firms combine relatively low powered incentives with a lot of control instruments and use hierarchy rather than courts to settle disputes. Markets are polar opposites, and hybrids are located in between.

As with Coase (1937), Williamson’s focus was on the contractual or administrative organization of production and its internal incentives. Likewise absent here – seventy years after Coase’s classic article – was any notion of the contracting entity that owns the means and fruits of that production, or can be sued if its outputs are defective. Although the research program has moved from the dichotomy to a continuum, these sizeable omissions have been thematic for Coase-Williamson-type transaction cost economics from the beginning. To their eternal credit, Coase and Williamson opened the black box of the firm to inspect its contents. But they forgot about the box itself.

Reinstatement of legal personality does not mean that we have to ditch the concept of transaction costs. Indeed, legal personality and legal incorporation are vital because they can greatly reduce such costs. When a firm is formed, the individuals involved do not have to contract and construct anew the structure of the firm. Instead, they can adopt the tried-and-tested legal template of a corporation, backed by the enforcement powers of the state (Robé 2011).

2. The myth of the firm-market hybrid

As well as within standard transaction cost economics, it became widely popular for other economists and sociologists to argue that the boundaries of the firm were fuzzy and indistinct. Ideas emerged of ‘internal markets’ within firms (Doeringer and Piore 1971), of the ‘quasifirm’ (Eccles 1981), and of ‘hybrid forms’ (Ménard 1995, 1996). Arguments went like this: the observed kind of contracting is typical of neither a market nor a firm, so it must be some kind of ‘hybrid’ of the two. For example, Robert Eccles (1981, pp. 339-40) considered ‘quasifirms’ in the construction industry where ‘relations between the general contractor and his subcontractors are stable and continuous over fairly long periods of time and only infrequently established through competitive bidding.’ He overlooked the possibility that these ‘stable and continuous’ contracting relationships may be neither markets nor (quasi) firms, but examples of relational exchange between different legal entities (including firms), along the lines of George B. Richardson (1972), Victor P. Goldberg (1980) and Ronald Dore (1983). Eccles (1981, pp. 342-3) went on to consider the divisions within the M-form (multi-divisional) firm:

These autonomous divisions function in many ways like independent firms. Often they compete successfully with firms in the market for business with each other. Thus internal transactions in these firms can have similar characteristics to market transactions. In

---

5 Doeringer and Piore (1971, pp. 1-2) admitted that ‘internal labor markets’ are not governed primarily by the price mechanism but by ‘a set of administrative rules and procedures.’ Marsden (1986, p. 162) went further: ‘internal labour markets offer quite different transaction arrangements, and there is some doubt as to whether they fulfil the role of markets.’
contrast, the inside contracting system is a set of market transactions with similar characteristics to hierarchical transactions. Most generally, pure markets and pure hierarchies are at opposite ends of a continuum of contracting modes.

In a sense it is possible that internal divisions within a firm might ‘compete’ with each other for business. But such competition would be internal point-scoring exercises rather than contractual or market competition. If a division wins a contract with an outside firm, then the legally enforceable agreement is not between that division and the outside firm, but between the firm that hosts the division and the outside firm. If separate divisions of a single firm were to compete fully for custom and contracts with a single outside company, then this would be largely self-defeating. It would be against the interests of the firm as a whole to allow its divisions to compete with each other on the price to be charged to an external customer. The key test is whether these divisions have separate legal status, and are recognized as legal persons. If they are not, then competition between divisions of the same firm is neither price-making nor market competition.

In a seminal article, Steven Cheung (1983, p. 11) also attacked the firm-market dichotomy, using this real-world example:

A landlord, who wants to build a high-rise finds a building contractor. This contractor subcontracts with a hardwood floor contractor on an agreed price per square foot – a piece count. The subcontractor, who imports the wood materials and adds finishing work to the wood on a piece-rate basis, in turn finds a sub-subcontractor, provides him wood, and offers him a price per square foot laid. Finally, the sub-subcontractor hires workers and again pays them per square foot laid.

Such complex integrations of contracts and subcontracts are very common. Cheung then saw elements of both a ‘firm’ and a ‘market’ in such an arrangement. Repeatedly stressing the payments per square foot, Cheung implied that piece-rate payments mean the existence of a ‘market’ (p. 10). But there is no reason why the one should imply the other. Employees within a firm can be paid by the piece or by the hour. Cheung (p. 17) also suggested that being ‘vertically integrated by contracts, with transfer pricing,’ might suggest that ‘only one firm exists.’ But being ‘vertically integrated by contracts’ is not the same as vertical integration within a firm. Putting two weak arguments together, Cheung suggested that the hardwood floor example is something with the characteristics of both ‘a ‘market’ and ‘one firm.’ The myth of the firm-market hybrid came of age.

Cheung’s hardwood floor example involved contractual exchange between at least five legal persons: the landlord, the contractor, the subcontractor, the sub-subcontracted firm, and its employees. The contracts had relational characteristics, and may have not involved market contracting (in the narrower sense defined previously). Considering another example, Cheung (1983, pp. 16-17) wrote:

If an apple orchard owner contracts with a beekeeper to pollinate his fruits, is the result one firm or two firms? The question has no clear answer. The contract involved may be a hive-rental contract, a wage contract, a contract sharing the apple yield, or, in principle, some combination of these and still other arrangements. In each case the beekeeper receives a remuneration for his service, and the orders he expects from the orchard owner vary with the form of contract.

In this case the answer to the question ‘is there one firm or two?’ depends on whether the contract involves hive rental, wage labour, or sharecropping. Cheung (p. 17) admitted this when he wrote: ‘Most economists would probably opt for only one firm if the beekeeper is
hired on a wage contract but for two if the hives are rented.’ This was a reasonable answer. The problem was not that he lacked a good answer, but that he was reluctant to endorse it. He was swept along by the rhetoric of the firm-market dichotomy.

Claude Ménard (1995, p. 176) tried to demonstrate that ‘organizations can be internally structured as quasi-markets’. Ménard considered franchising ‘when very strict standards are imposed on independent participants.’ He noted:

Classification [into markets or organizations] becomes particularly difficult when firms are interconnected by a dense web of transactions, with strong commitments to each other and complementarities of their assets, but without formal agreements and, moreover, with property rights on these firms clearly maintained as distinct.

He thus proposed intermediate ‘forms’ between markets and hierarchies. But instead of focusing on the ‘strong commitments to each other’ to then describe them as relational exchange, he instead saw them as ‘hybrids’ involving ‘specific combinations of markets incentives and modalities of coordination involving some form of hierarchical relationship’ (p. 175).

In fact, Ménard’s example of strictly monitored and regulated franchising is an ‘organized’ relationship between two or more firms or legal persons. Although the relationship has an organized or even hierarchical character, that does not mean that it constitutes a single firm. Many contracts between separate legal persons involve ties that allocate roles and responsibilities. This does not necessarily mean that they become a single legal entity. The fact that ‘property rights on these firms [are] clearly maintained as distinct’ underlines the fact that multiple firms may be organized, just as individuals can be organized, without necessarily creating a singular legal person or a firm. Ménard’s error was to confuse broadly-defined organizational arrangements with the firm. He also overlooked the fact that relational contracting involves different firms, but because of its relational character, it is not on a market. It is thus a third option, after a market and a firm. Relational contracting means that the firm-market dichotomy is false. It also undermines ‘hybrids.’ Illusions of a ‘firm-market hybrid’ disappear once we adopt the notion of the firm as a singular legal entity.

3. Some other influential views on the firm

Armen A. Alchian and Harold Demsetz (1972) famously conflated the firm with the market. They presented ‘the firm and the ordinary market as competing types of markets’ (p. 795), and argued that market transactions were not eliminated within the firm. They declared that ‘the firm … has no power of fiat, no authority, no disciplinary action any different in the slightest degree from ordinary market contracting between any two people’ (p. 777). In their view, an employer has no more authority over an employee than a customer has over a grocer.

For Alchian and Demsetz the principal difference between intra-firm relationships (a grocer and his employee) and ordinary market contracting (a grocer and his customer), lay in the ‘team use of inputs and a centralized position of some party in the contractual arrangements of all other inputs.’ Their challenge was to explain why, ‘instead of multilateral contracts among all the joint inputs’ owners, a central common party to a set of bilateral contracts facilitates efficient organization of the joint inputs in team production’ (p. 794). With team production the overall output is observable, but each individual’s contribution to the output is difficult to determine. This creates an ‘incentive to shirk’ or free-ride (p. 780). Team members realize that their effort can be reduced without a proportional loss of their own income.
Hence according to Alchian and Demsetz (1972, pp. 781-2), it made sense for ‘someone to specialize as a monitor to check the input performance of team members,’ provided that the monitor has the ‘incentive not to shirk as a monitor.’ This incentive is to attribute ‘residual claimant status,’ that is ‘title to the net earnings of the team, net of payment to other inputs,’ to the monitor.

This explanation of the nature and existence of the firm differed radically from that of Coase or Williamson, particularly by its omission of transaction costs. With Alchian and Demsetz (1972) the firm is simply the individual residual claimant who monitors the team. He or she alone reaps the profits, may sue suppliers, or be sued by customers. The firm can only be a self-employed contractor. If an attempt were made to include organizations of multiple human individuals as firms, then the alleged monitoring problem would re-emerge, and one of these individuals – according to their logic – would have to become the residual claimant. The multi-individual firm would then revert back into a single-individual entity. Despite the insights in this paper, it cannot deal adequately with multi-person firms or corporations.

In a highly influential article, Michael C. Jensen and William H. Meckling (1976) developed their ‘nexus of contracts’ view of the firm. Much of the article involves formal modelling, but its opening pages contain an important conceptual statement. At the outset the authors identified a major defect in the approach of Coase and Williamson – the common failure to take into account contractual relations with customers for the firm’s output. Jensen and Meckling (1976, p. 310) wrote: ‘Contractual relations are the essence of the firm, not only with employees but with suppliers, customers, creditors, and so on.’ They took contracts with customers into account and also stressed the role of law. In explicit terms, Jensen and Meckling (1976, p. 311 n.) acknowledged ‘the important role which the legal system and the law play in social organizations, especially, the organization of economic activity.’ They continued:

Statutory laws sets [sic] bounds on the kind of contracts into which individuals and organizations may enter without risking criminal prosecution. The police powers of the state are available and used to enforce performance of contracts or to enforce the collection of damages for non-performance.

This is a rare acknowledgement of the economic and real significance of the legal aspect of the firm. Jensen and Meckling (1976, p. 311) admitted that the firm can be a contracting agent. They wrote of ‘contracts … between the legal fiction (the firm) and the owners of labour, material and capital inputs and the consumers of output.’ Likewise, in another article, Jensen and Meckling (1983, p. 9) explained: ‘Individuals and organizations – employees, investors, suppliers, customers, etc. – contract with each other in the name of a fictional entity – the corporation.’ Here again it was suggested that ‘organizations’ as well as individuals can contract, and ‘in the name of’ the corporation. This seemed to admit the possibility of corporate legal personality. But ultimately this was denied.

Although Jensen and Meckling noted ‘the artificial construct under the law which allows certain organizations to be treated as individuals’ (p. 310 n.), they argued that this is simply a ‘legal fiction which serves as a nexus for contractual relationships’ (p. 311). ‘Viewing the firm as the nexus of a set of contracting relationships among individuals,’ Jensen and

---

6 In the citation-based study by Kim et al. (2006) of ‘what has mattered’ in economics since 1970, Jensen and Meckling’s (1976) article features in third position. The Alchian and Demsetz (1972) article is in twelfth place.
Meckling (p. 311) warned that ‘the personalization of the firm … is seriously misleading. *The firm is not an individual.*’ As Jensen (1983, p. 327) elaborated elsewhere:

*The nexus of contracts view of organizations helps to dispel the tendency to treat organizations as if they were persons. Organizations do not have preferences, and they do not choose in the conscious and rational sense that we attribute to people … the behavior of the organization is like the equilibrium behavior of the market … the individual agent is the elementary unit of analysis.*

There are several points of contention here. One concerns the meaning of a ‘legal fiction.’ Another concerns the claim that organizations cannot be treated as agents, contrary to many other writers.

Addressing the latter point first, of course organizations do not have their own separate minds or brains. But whether organizations have preferences, or are rational, depends on the adopted meanings of those terms. The variability of behaviour that Jensen and Meckling observed in organizations is also found in individuals. We can procrastinate or change our minds. There is also the hypothesis of ‘multiple selves’ (Elster 1986). The differences between individuals and organizations are real, but not entirely as Jensen and Meckling claimed. At root of their argument was a refusal to admit that, when there are clear procedures for resolving internal conflicts and making decisions, organizations can be treated for some purposes as singular agents. This does not imply that individuals are unimportant or dispensable, or that organizations are themselves conscious entities.7

The position of Jensen and Meckling on ‘legal fictions’ was also problematic. *Legal fictions are not false* (Fuller 1967). They are devices used in legal reasoning to transfer principles that have been established in one context to another. For example, once an order of child adoption is entered, the biological parents become legally unrelated to the child and lose their parental rights. The adoptive parents are legally considered to be the parents of the adopted child. A new birth certificate may be issued. It is a legal fiction. But, as with other legal fictions, it is backed by the power of the law. It is real.

In the case of the corporation, the legal fiction involves a transfer of rights and liabilities concerning ownership and contracting from individuals to registered corporate organizations. This does not necessarily mean that the law grants corporations all the rights that it grants to individuals. Important here are rights to own assets, to enter into contracts, to sue, and be sued. But other legalities are different: people who terminate a corporation are not charged with murder.

Jensen and Meckling (1976) failed to acknowledge these nuances. Viewing the firm simply as ‘a set of contracting relationships among individuals’ (p. 311) they failed to show how this ‘nexus’ itself forms a contract with suppliers or customers if it is no longer deemed to be a singular legal entity. They did not consider how the problem of how to deal with the death, bankruptcy, or insanity of one of the individuals making up the nexus. In short, they evaded the issue of how the firm survives the legally-operational lives of the individuals in the nexus. Corporations themselves can face critical life and death events, such as the merger, acquisition, or liquidation of the entity. These differ from simple contractual relations between individuals.

---

7 Coleman (1982), North et al. (2009) and List and Pettit (2011) all argue that under restrictive conditions organizations can be treated as agents.
The key point of contention is whether an organization, as well as a human individual, can itself enter into contracts and have ownership rights. Terms such as *legal person*, *legal personality* or *legal fiction* should not mislead us into thinking that anything else of major importance is necessarily at stake here. When the law sees the firm as a person, it means no more than that the law treats the firm as a point of imputation for legal rights and duties. Legal ‘personality’ involves an analogy. In his discussion of corporate personality, John Dewey (1926, p. 656) pointed out that when we describe a wine as ‘dry’ we do not mean that it is no longer a liquid – we have merely followed conventional usage and taken this word from one context and applied it to another.

Jensen and Meckling (1976, p. 311) also blurred the distinction between a firm and a market. They wrote:

> The firm is … a legal fiction which serves as a focus for a complex process in which the conflicting objectives of individuals … are brought into equilibrium within a framework of contractual relations. In this sense the ‘behavior’ of the firm is like the behavior of a market, that is, the outcome of a complex equilibrium process.

On the contrary, a corporation need not be in equilibrium. All that is required of it is the capacity to establish a coherent and legally recognizable position, through appropriate legal procedures, when making any contract with another legal entity. Although bound by its past contracts, organizations like humans can change policies back and forth. They can – and sometimes do – develop incoherent or unstable strategies. Although individual contracts may lead to the formation of a firm – as with a partnership contract – they are not contracts to which the firm itself is a party. Negotiations and votes, that take place at meetings of boards of directors or shareholders, are neither markets nor legal contracts. They are decision-making processes within an already-established framework of legal and internal rules. The claim that the firm is like a market is illogical and ungrounded.

Jensen and Meckling (1983, p. 10) hinted at a different source for their anxiety and an alternative account of the origin of the corporation:

> the corporation is *neither* the creature of the state *nor* the object of special privileges extended by the state … the corporation requires for its existence only freedom of contract. Corporate vitality is in no way dependent on special dispensation from the authorities. Limited liability, for example, is not an idea specialized to corporations. Non-profit organizations, partnerships, and individual proprietorships, for example, all exhibit various forms of limited liability … Freedom of contract surely encompasses the right of parties to prescribe limits to liability in contracts.

Their argument here was not against corporate legal personality as such, but against the idea that corporate vitality and limited liability are necessarily creations of the state. They suggested that these arrangements can emerge spontaneously. All that is required is ‘freedom of contract.’

The final case study in this section is the ‘new property rights theory’ of Sanford J. Grossman, Oliver D. Hart, and John H. Moore (Grossman and Hart 1986, Hart 1989, 1995, Hart and Moore 1990, Moore 1992). This approach builds on several preceding theories but is critical of their limitations. Hart (1989, p. 1764) argued that ‘the nexus of contracts approach does less to resolve the questions of what a firm is than to shift the terms of the debate.’ Moore (1992, p. 494 n.) wrote similarly: ‘One can … sidestep the issue entirely, by arguing that everything is contractual, and that firms are a mirage … This is the view proposed by
Jensen and Meckling … But if firms are a mirage, it is difficult to explain the enormous resources that firms expend merging and breaking up.’

Grossman and Hart (1986, pp. 692-3) ‘define the firm as being composed of the assets (e.g., machines, inventories) that it owns. … We define a firm to consist of those assets that it owns or over which it has control.’ Or, more cruelly, the firm was viewed ‘as a collection of physical assets’ (Hart and Moore 1990, p. 1121). But a pile of physical assets is insufficient to constitute a firm. Ironically, by focusing solely on control, they eviscerated the term ‘property rights’ that they used to describe their theory. In general an entity cannot be defined solely in terms of what it controls.

Grossman and Hart (1986, p. 693) saw a major problem in previous theories as the lack of ‘a sufficiently clear definition of integration.’ They were right in their identification of a key problem, but deficient in terms of its solution. They explained integration in terms of owned non-human assets, including machines, buildings, contracts, and patents. These allegedly hold firms together and explain how their managements exert effective powers of authority over employees. In their theory ‘the firm’ owns these assets and holds rights of control over them as well as the rights to the residual income of the enterprise. Workers require use of those assets, and may have few alternative options, especially if they require specialist equipment. Consequently, ‘authority over assets translates into authority over people’ (Hart and Moore 1990, p. 1150). As Hart (1995, pp. 57-59) argued:

A firm’s nonhuman assets … simply represent the glue that keeps the firm together … If such assets do not exist, then it is not clear what keeps the firm together … One would expect firms without at least some significant nonhuman assets to be flimsy and unstable entities, constantly subject to the possibility of break-up or dissolution.

This formulation is inadequate, on at least two counts. First, a collection of assets cannot ‘simply represent the glue that keeps the firm together.’ Even in their own terms it is ownership or control of these collections of assets that provides the firm with power over employees and suppliers. These powers are backed by law. As Scott Masten (1991, p. 208) insisted: ‘Ownership itself is a condition sustained by legal rules and remedies.’ Yet Grossman, Hart and Moore paid even less attention to the legal nature of the firm than some of the competing theorists. Hart (2011, p. 102) asked: ‘is a firm circumscribed by its legal status or by its economic activities?’ This is a false and misleading dichotomy. The ‘economic’ activities of the firm become possible because the firm has a legal status, and has powers enshrined in law. As Simon Deakin (2006, 2012) explained, legally recognized capacities define conditions of access to the market; the capacity to own assets and enter contractual relations. As Edward M. Iacobucci and George G. Triantis (2007, p. 518) put it:

Legal persons may vindicate their ownership rights in court, and they may be defendants against whose property creditors may enforce their claims. Accordingly, only a legal person has the capacity to contract – that is, to make a legally enforceable pledge of its assets to the performance of its promise.

A collection of assets without a legal person as their owner is no more a firm than a collection of bones, flesh, and blood is a human being. Once the legal and ownership aspects of the problem are fully acknowledged, we face the question of how the firm establishes itself as a unitary legal entity, and becomes more than a collection of contracting individuals. Bengt Holmström (1999, p. 87) thus wrote: ‘Individual ownership of assets does not offer a theory of organizational identities unless one associates individuals with firms.’ Holmström (p. 100) concluded: ‘property rights theory, as articulated in Hart and Moore (1990) and other
representative pieces, says very little about the firm. The problem is that there are really no firms in these models, just representative entrepreneurs.’

If non-human assets are insufficient to keep the firm together, are they necessary? Raghuram G. Rajan and Luigi Zingales (2000, 2001) argued that they are not, pointing out that many modern firms are highly labour and knowledge intensive. This puts more bargaining power in the hands of the highly-skilled knowledge worker and relatively less in the hands of the owner of physical assets (Hodgson 1999). According to this argument, the relative importance of ownership of physical assets is diminishing in modern capitalism. There are knowledge-intensive firms such as software producers that rely on relatively inexpensive hardware. And there are so-called hollow corporations – such as Benetton or Marks and Spencer – that subcontract out much of their manufacturing; their corporate identity and competences are more to do with image and marketing than material assets.

These puzzles concerning the nature and identity of the firm are solved once we treat it as a legal entity. The glue that holds the firm together consists of the contracts or articles that bind the parties into one legal entity, with its own legal rights and obligations. The organization becomes a firm when it acquires a legal personality; its ownership of assets is secondary. In the case of the corporation, the ‘glue’ can outlast the lifetimes of the individual members involved. Some corporations are hundreds of years old. Their individuals and assets have changed many times over, but the corporations live on.

Crucially, the firm is distinct from any or all of its human constituents, contrary to views of the firm as a coalition of owners, and it is distinct from any and all of its non-human assets, contrary to the ‘new property rights theory’ of Grossman, Hart and Moore. A coalition of owners may create a firm. Firms typically own non-human assets. But the firm is not the same thing as a coalition of individuals or a collection of assets.

There is no good reason for economists to relinquish a legally-grounded definition of the firm or corporation. All major theories of the firm depend on legal concepts – particularly ownership – despite the neglect of the firm itself as a legal entity. Legal specifications and frameworks are vital for the firm to operate. Legal relations are an unavoidable part of the definition, alongside other factors. This argument is strengthened, not undermined, by the real growth of other economically significant entities such as business units, conglomerates, keiretsu, strategic alliances, supplier networks, relational contracting, and so on. Each of these entities makes use of legal forms, including contracts and property rights. Indeed, the growth of a diversity of business and industrial structures makes it imperative to develop clear, distinct definitions of the different entities involved and to understand their legal structure. A muddled reality is no excuse for muddled definitions. Likewise, a mutable reality is no justification for elastic ideas. To describe or understand a tangled reality we need clear concepts and careful definitions to guide us.

4. Conclusion: defining the firm

Although firm is used here in the broad sense of a legally recognized unit set up for the purposes of producing and selling goods or services, there are other productive formations that are different from firms and also have to be defined. For example, P. Sargant Florence (1957) considered matters of control and possible oligopolistic concentrations of power.

---

Florence (p. 244) argued: ‘To economists, more directly interested in degrees of competition and monopoly than in industrial location, the firm, the unit of control, is more important than the plant, the physical unit.’ In this context, he placed more emphasis on the ‘unit of control’ than the legally defined entity of the firm. The problem was how to deal statistically with subsidiary companies that legally were separate firms but were managed and controlled by their parent companies. Florence endorsed the practice of the UK Census of Production, which used the term ‘business unit’ to refer to a parent company together with any subsidiary companies of which it owned more than 50 percent. Hence a ‘business unit’ was not the same as ‘a single company in the legal sense.’ Neither would it include subsidiaries in which the interest of the parent company was in practice sufficient for managerial control, but did not amount to as much as 50 percent. Florence made it clear that firms and business units were different things and required different definitions.

Fritz Machlup (1967, p. 26) argued differently. He listed no less than ten ‘concepts of the firm employed in the literature of business and economics’ and suggested that even more were in use. He argued that the firm had been regarded as an organization, a decision-making system, a collection of assets and liabilities, a juridical person, a business unit under a single management, and much more besides. Machlup (1967, p. 28) then concluded:

This exercise should have succeeded in showing how ludicrous the efforts of some writers are to attempt one definition of the firm in economic analysis … I hope that there will be no argument about which concept of the firm is the most important or the most useful. Since they serve different purposes, such an argument would be pointless.

But clearly this is a non sequitur. The fact that several different conceptions of the firm exist does not imply that the formulation and promotion of one best definition should be abandoned. On the contrary, it could be argued that the very confusion itself calls out for a single, commonly accepted definition. Machlup rightly pointed out that these different concepts ‘serve different purposes.’ But it did not occur to him that these different definitions might also point to different things. A (legal) firm is not the same thing as a business unit, or a collection of assets, or a production plant. These are not simply different concepts: they are different entities. Accordingly, we require a plurality of concepts to refer to a plurality of real arrangements.

Cheung (1983, p. 17) argued that any definition of the firm is subjective and arbitrary: ‘according to one’s view a ‘firm’ may be as small as a relationship between two input owners or, if the chain of contracts is allowed to spread, as big as the whole economy.’ He then concluded: ‘Thus it is futile to press the issue of what is or is not a firm’ (p. 18). Cheung’s argument was basically this: because different definitions of $X$ are possible, it is futile to define $X$. But this too is a non sequitur. If different definitions exist, then to communicate we have to make clear the one we have chosen rather than abandon definitions altogether.

No good argument has been presented to abandon a legal conception of the firm. It is also important to adopt additional concepts such as business unit, production plant, and conglomerate, to describe other important structures in the business world. Multiple concepts are required to describe multiple real entities. Instead, it is repeatedly proclaimed that all attempts to define the firm are fruitless. This licensed imprecision promotes habits of conceptual vagueness and terminological sloppiness that have marred the literatures in economics and business studies ever since. It is no wonder that progress in the theory of the firm has virtually ground to a halt.

Claims that the boundaries of the firm have broken down, and that there is no essential distinction between firms and markets, are misconceived. Internal markets within firms prove
to be chimerical. Alleged cases of firm-market ‘hybrids’ turn out to be interlocking relations or networks between multiple and distinct legal firms or legal persons, rather than an encompassing firm or hybrid form.

Recognition of the absence of markets and commodity exchange inside firms is important for several reasons. It dispenses with confused terms such as ‘internal market’, ‘continuum’ and ‘hybrid.’ It also helps underline the relevance of the boundaries of the firm and the vital interface between non-market and market modes of co-ordination. Any analysis of the formation and role of these boundaries has vital implications for corporate and public policy. A neglect of legal realities impairs any attempt by the social scientist to give advice on appropriate legal structures to enhance economic performance.

Furthermore, without attention to legal relations, social scientists are ill-equipped to intervene in the long debate concerning the limitation of abuses of corporate power. They will be less able to evaluate the conditions involved in any legal incorporation of a firm by the state, and the nature of the quid pro quo for society in return for the legal privilege of limited liability. Without attention to these features, social scientists may become dangerously indifferent to policies that extend or diminish the scope of corporate power or the real market. They will be unable to engage effectively in important debates concerning corporate law reform, and the development of corporate structures that are more conducive to social welfare.

The firm has at least two fundamental features: (1) it is set up with resources and capabilities to produce goods or services for sale, and (2) in owning assets, contracting inputs and selling outputs it acts as a single legal person. As a legal person, the firm has legal ownership of the goods as property up to the point that they are exchanged with the customer, the legal right to obtain contracted remuneration for the produced services, and the potential liability to be sued for non-fulfilment of contracts with suppliers or customers.

What do we mean by the ‘boundaries’ of the firm? There is something in the argument of Jensen and Meckling that the firm is more a nexus than a bounded zone. But condition (1) sharpens this nexus from being any contracting agent or individual, to give the firm the distinct purpose of production for exchange. This makes it more than a nexus: it is a functional entity. Furthermore, the vital stipulation of legal personality in condition (2) constitutes the nexus and provides the ‘missing person’ that is absent in much of the theory of the firm (Gindis 2013).

An individual trader is a nexus of contracts, but has mind, body, memory, expectations, and intentions as well. A firm is also a nexus of contracts but similarly is more than that. All firms have the capacity to produce and sell a product. When the firm is an organization it also has structure, managers, routines, capabilities, strategies, and much more.

The principal structural bonds that keep the multi-person firm together are fixed by the founding agreement of association, to be altered infrequently, rather than through everyday renegotiation. These bonds are very different from market relationships. The firm is a distinct legal entity. It owns its products and sells or hires them to others. It enters into contracts with its workforce and its customers. Accordingly, its external relations involve contracts and sometimes markets. But neither markets nor exchange can exist inside the single legal entity of the firm. To engage in production, firms manage their own or hired resources by administrative control rather than via markets and exchange.

Coase (1937) started a long debate on the nature of the firm that is still unresolved. The argument here is that theorists have basically been looking in the wrong direction. They have looked to managing and contracting in production as the portal to an understanding of the
nature of the firm. Instead, the firm is constituted by its relations with society more broadly, including the state. External relations and legal powers provide the glue that holds the firm together. They may bind partners together with durable contracts, or they may constitute the firm as a corporate entity. The firm is an institution made not by private ordering, but through the interactions of business with the established rules of contract and company law. Even if the legal notion of a firm tells us little of how production is or can be organized, it is vital to understand how a firm interacts with suppliers and customers.

References


