Abstract

In the last four decades, one of the fastest-growing fields of research in economics has been the contractual theory of the firm developed in Coase’s (1937) footsteps. Yet despite what otherwise seems to be a genuine success story the question of the nature of the firm remains an empirical and theoretical challenge, painfully illustrated by the lack of consensus regarding the definition and boundaries of the firm. The argument of this thesis is that many thorny questions that plague the literature, including issues related to ownership, boundaries, and intra-firm authority, are due to the fact that contractual theorists of the firm have generally overlooked a key legal feature of the economic system, without which theories of the firm are like Hamlet without the Prince.

An elementary institutional fact about firms and markets is that in order to become a fully operational firm in a modern market economy, an entrepreneur or an association of resource owners need to go through a registration or incorporation procedure by which the legal system creates a separate legal person or legal entity in which ownership rights over assets used in production are vested, in whose name contracts are made, and thanks to which the firm has standing in court. With this assignment of legal personality, the legal system creates the efficiency-enhancing nexus for contracts that literally carries the organizational framework of the firm, and secures its continuity by locking-in the founders’ committed capital, thereby
allowing them to pledge assets, raise finance and do business in the firm’s own name.

Given the basic principle that only legal persons may own property and have the capacity to contract, and the implication that legally enforceable contracts can only exist between legal persons, it is something of a paradox that the notion of legal personality is absent from the prevailing narrative in the contractual theory of the firm. The thesis examines the reasons behind this state of affairs, and identifies alongside the widespread view among economists that firms can be defined with little or no reference to law, particularly statutory law, the lasting influence of Jensen and Meckling’s (1976) ambiguous dismissal of legal personality as a legal fiction that unavoidably leads to misleading reification.

In order to disentangle the issues involved, the thesis puts this argument into historical perspective, and suggests that much can be learned from the corporate personality controversy that in the past has addressed the same questions. As the overview of the history of this debate reveals, the category mistakes that Jensen and Meckling presented as inevitable can be easily avoided once the meaning and functions of legal personality are properly understood. The thesis dispels enduring misunderstandings surrounding the notion of personhood, and proposes a legally-grounded view of the nature and boundaries of the firm that recognizes in law’s provision of legal entity status a fundamental institutional support for the firm while fitting the overall Coasean narrative.
For Nina
Things fall apart; the centre cannot hold.

Yeats (1934 [1920]: 211)
Acknowledgements

This thesis is the result of a long and tortuous journey that owes much to the extraordinary patience of my supervisor, Professor Geoffrey Hodgson, whose work has inspired and guided me over the years.

Many other scholars and friends have helped me frame the questions and understand at least in part the answers. For these contributions made at seminars, conferences, by email, or over a glass of wine, my thanks go to Richard Adelstein, Richard Arena, Bernard Baudry, Laure Bazzoli, Yuri Biondi, Béatrice Boulu-Reschef, Robert Boyer, Olivier Brette, David Copp, John Davis, Jan Dejnožka, Michael Dietrich, Pierre Dockès, Denise Dollimore, Véronique Dutraive, Dave Elder-Vass, Olivier Favereau, Pierre Garrouste, Pierre-Yves Gomez, Xavier Hollandts, Antoine Jeammaud, Thierry Kirat, Alex Lascaux, Nathalie Lazaric, Jan-Willem Lindemans, Pavel Luksha, Harro Maas, Uskali Mäki, Steven Medema, Ngaire Naffine, John O’Neill, Eric Orts, Huascar Pessali, Chris Renwick, Jean-Philippe Robé, Jan-Willem Stoelhorst, and Bertrand Valiorgue. I am also very grateful to my fellow long-distance runner, Benjamin Chapas, with whom many topics in economics and philosophy were discussed at length.

Since my arrival at the Hertfordshire Business School I have received valuable support and advice from Frank Currie, Hulya Dagdeviren, Jon Easter, Jane Hardy, and Keith Randle. I would also like to thank Angela
Digby, whose help with various administrative tasks has gone beyond the call of duty.

On a more personal level I am indebted to Lydie Chaurand and Serge Caranzan who encouraged me in various ways. My thoughts are also with my parents, Nina and Isaak, and my brother, Daniel, for whom the good days lie ahead. Last but not least, this work would not have been completed without the incredible support of my wife and colleague, Francesca Gagliardi.
# Contents

*Acknowledgements*  
i  
*Contents*  
iii  

## Chapter 1. The firm as a challenge  
1  
1.1 Success and failure in the theory of the firm  
2  
1.2 Aims and scope of the thesis  
8  
1.3 Structure of the thesis  
14  

## Chapter 2. From dichotomy to continuum  
19  
2.1 Firms versus markets  
20  
2.2 Firms as markets  
35  
2.3 Lost in the continuum  
47  

## Chapter 3. Reclaiming boundaries  
57  
3.1 The fundamental transformation  
59  
3.2 The meaning of ownership  
72  
3.3 In search of new foundations  
84  

## Chapter 4. The nexus paradox  
97  
4.1 The definitional impasse  
99  
4.2 Missing persons  
114
4.3 The reification illusion 128

Chapter 5. The corporate personality controversy 143

5.1 From status to contract 145

5.2 Neither contract nor concession 157

5.3 The value of corporate personality 170

Chapter 6. Hamlet with the Prince 185

6.1 The meaning of personhood 187

6.2 The firm, the market and the law 199

6.3 Future research 209

Cases cited 213

References 214
Chapter 1. The firm as a challenge

What is a firm? Since … the beginning of the 1970s much progress has been made, yet despite the important literature on the subject, this question remains an empirical as well as theoretical challenge.

Garrouste and Saussier (2005: 179)

It is customary to begin a discussion of the theory of the firm developed since the 1970s with the observation that the “black box” of standard neoclassical economics had finally been opened. At long last, the argument goes, the textbook depiction of the profit-maximizing firm as a set of production possibilities, which simply assumed that firms exist and had no need to consider internal organization, had given way to a new line of inquiry that sought to explain how and why firms come about in the first place, how they are organized, and what determines the optimal perimeter of their activities. The theory of the firm developed to answer these questions has been described not only as “one of the fastest-growing areas in applied
microeconomics” (Klein, 1999: 463) but also as a “big business” (Gibbons, 2005: 200). Not surprisingly, all of this literature’s foundational articles are highly cited and feature in Kim, Morse and Zingales’s (2006) ranking of “what has mattered” in economics since 1970.

Significantly, in addition to the Nobel Prize in Economic Sciences attributed to Coase in 1991, Williamson, the other towering figure of the field, shared the Nobel Prize in 2009 for his work on the boundaries of the firm that has led to further theoretical refinements by followers and critics alike, as well as to an impressive amount of empirical work. The overall result, according to Masten’s (2002: 432) discussion of modern evidence on the firm, is that “much more is known about the determinants of organizational form now” than ever before. However, despite what otherwise seems to be a genuine success story the question of the nature of the firm remains an “empirical as well as a theoretical challenge,” as Garrouste and Saussier (2005: 178) were recently obliged to concede. This introductory chapter proposes an overview of the success and failure in the theory of the firm literature (1.1), before presenting the aims and scope of the thesis (1.2), and summarizing its structure (1.3).

1.1 Success and failure in the theory of the firm

In his landmark work, Markets and Hierarchies, Williamson (1975: 20) claimed that “in the beginning there were markets.” This heuristic dictum captures with a Biblical ring the widespread view of the agenda of an economic theory of institutions: start, in an institution-free world, viewed as
a sort of Hobbesian “state of nature,” with self-interested human beings possessing a “propensity to exchange” (Smith, 1937 [1776]: 13), and attempt to explain the endogenous emergence of the key institutions of social order, such as money, property, contract, law, courts, firms, and states. Whether developed in the game-theoretic language of Schotter’s (1981) *Economic Theory of Social Institutions*, or in the non-formalized style of Williamson’s (1985a) *Economic Institutions of Capitalism* and Barzel’s (2002) *A Theory of the State*, research carried out according to this agenda was described by Williamson (1975: 1) as the “new institutional economics.”

One of the ideas successfully promoted by Coase (1960) and other early new institutional economists is that institutions of various kinds are necessarily irrelevant in the zero transaction cost world of conventional price theory. In this “strange world” (Furubotn and Richter, 1998: 8) of perfect information and costless exchange, bargaining always ensures allocative efficiency, irrespective of the institutional setting within which exchanges take place. Neither initial allocations of property rights, nor specific types of liability rules, nor indeed the structure of law itself, make a difference to the final outcome. By contrast, according to the Coasean narrative, the real world does not fit into this neat picture. Positive information and transaction costs are pervasive, and may sometimes be sufficiently high to outweigh the

---

1 According to Klein (1999: 456), new institutional economics is “an interdisciplinary enterprise” that “borrows liberally from various social-science disciplines” in order to “understand the institutions of social, political and commercial life” while retaining economics as “its primary language.” See Ménard and Shirley (2005), and Brousseau and Glachant (2008), for recent comprehensive overviews of this tradition.
benefits of certain transactions, thus preventing allocative efficiency. This means that transaction cost-reducing institutions of various kinds can be efficiency-enhancing, and that the comparative economic analysis of alternative institutional arrangements is justified.

The contractual theory of the firm plays a prominent role within the new institutional economics given that “perhaps the most important adaptation to the existence of transaction costs is the emergence of the firm” (Coase, 1988a: 7). Unlike the firm of conventional price theory, the efficiency of the contractually-viewed firm is not a question of production possibility frontiers and optimal responses to exogenous price and cost variations. To the extent that combining the resources of several owners for production always “involves a partial or outright transfer of property rights through a contract” (Cheung, 1970: 50), efficiency implies that a set of agreements that no one will wish to change is somehow reached. Joint output value cannot be maximized without such a set of agreements.\(^2\) For the “contractual man” described by Williamson (1985a: 43ff), the efficient “governance of contractual relations” (Williamson, 1979a) implies designing institutional arrangements that improve coordination by aligning incentives and reducing the risks involved, thereby reducing the costs of exchange.

The fact that one observes several types of associations or organizational forms implies competition between various forms of contractual arrangements between resource owners. As Alchian (1984: 47, emphasis in original) put it, “competition occurs in forms of organization and

\(^2\) Such a set of agreements corresponds to a “nexus of strategies” played by rational agents (Dow, 2004: 525).
contracts.” Those forms that pass the “market survival test,” viewed along the lines of Alchian’s (1950) “natural selection” argument, reduce or minimize costs not only compared to stand-alone producers but also compared to other conceivable forms of organization (Jensen, 1983: 331). In the long run, in other words, the observed distribution between typical economic forms of the firm (e.g., capitalist or cooperative) and typical legal forms of the firm (e.g., partnership or corporation) is the result of competitive market forces. On this view, the forms best suited to a large variety of needs and circumstances tend to perform better than alternatives, and through a process of “propagation by imitation” (Demsetz, 1996: 489) tend to ultimately survive.  

One such form of contractual arrangement involves the delegation by resource owners “to a central agent, for some period of time, specific rights to direct their assets in production in return for a payment” (Eggertsson, 1990: 48; see also Pejovich, 1990: 53). The general consensus is that in situations “where economic agents cooperate with one another not through a system of explicit contracts that bind each to every other member of the group but through a system of bilateral contracts in which each comes to an agreement with a ‘single contractual agent’, the essential ingredient of ‘the firm’ is present” (Ricketts, 2002: 3). This is the “nature of the firm”

---

3 New institutional economists generally agree with Jensen (1983: 322) that “as a result of the subtle interaction of the continual striving by purposeful individuals and the natural selection properties of the environment, extremely complicated and sophisticated institutions and practices can arise.” See Vromen (1995: 56ff) for a detailed discussion of the new institutional economists’ conception of the evolution of efficient organizational forms.
originally theorized by Coase (1937), that later grew into Alchian and Demsetz’s (1972) depiction of the “classical capitalist firm,” from which follows Jensen and Meckling’s (1976) famous definition of the firm as a “nexus of contracts.”4 All contractual theories of the firm are explicit or implicit variations on this pivotal idea.5

The Coasean theory of the firm is based not only on the presence of this efficiency-enhancing central contractual agent but also on the claim that the system of contractual relations thus organized creates an “island of conscious power” (Coase, 1937: 388) that distinguishes the firm from ordinary market contracting, and can help explain the distribution of activities between firms and markets. Taking the first claim for granted, contractual theorists have concentrated mainly on the second claim, and debates have accordingly focused on the distinguishing mark of the firm. More precisely, the bulk of the literature has been concerned with the related question of the determinants of firm boundaries. Yet despite four decades of work on a clearly identified set of questions, there is neither a “unique” account of firm boundaries nor a “unified” theory of the firm (Garrouste and Saussier, 2008: 36). Instead, “many competing theoretical frameworks coexist, with only

---

4 Although there is a continuity between Coase’s question (“why firms?”), Alchian and Demsetz’s question (“why capitalist firms?”), and Jensen and Meckling’s question (“why corporations?”), these are not exactly the same questions.

5 Describing these developments, Furubotn and Richter (1998: 326) observed that “the firm is now viewed as a nexus of contracts that regulate the nonmarket transactions between resource owners who form ... a private enterprise.” This view has become so widespread that certain introductory texts (e.g., Wittman, 2006: 332) simply state that the firm is a nexus of contracts as a matter of fact, without further specification or explanation.
partial answers concerning the nature of the firm, its boundaries, and its internal organization” (Garrouste and Saussier, 2005: 194).

In fact, after substantial advances in the 1970s and 1980s, progress has considerably slowed down since the early 1990s, and in the last decade or so there has been no significant progress at all. Unfortunately, Eggertsson’s (1990: 158) observation, that the lack of standardized vocabulary and careful definitions “makes it difficult to see whether we are dealing with overlapping or competing theories,” still rings true today. In this context, Gibbons’s (2005: 227ff) effort to nest several existing theories of firm boundaries within a formalized “integrative framework,” the primary functions of which are “to differentiate among the theories by clarifying their distinctions,” and to allow the progressive integration of additional contractual explanations, is a significant step forward. However, although the integration of existing approaches is important, progress in the field is likely to involve “breaking loose” (Gibbons, 2005: 238), at least partially, from the contractual paradigm. Some sort of “new foundations” (Zingales, 2000) are needed.

The view that progress can be made if a previously neglected element is taken into account is hardly novel, and has been expressed by some of the leading theorists of the firm at least since the late 1980s. Coase (1988d: 38) himself had complained that a regrettable consequence of the “undue emphasis … on the choice of contractual arrangements” is that “economists have tended to neglect the main activity of a firm, running a business.” More

---

6 As if to illustrate this point the editors of the second edition of the *New Palgrave Dictionary of Economics* chose to reproduce the entry on the theory of the firm (Archibald, 2008) from the previous edition (Archibald, 1987).
research was needed, Demsetz (1988a: 157) had similarly argued, on the production, maintenance and use of knowledge, while for Milgrom and Roberts (1988: 450) “evolutionary considerations” related to changing technology deserved significantly more attention. These suggestions were welcomed by economists working outside the contractual tradition (e.g., Winter, 1988; Langlois, 1988; Teece, 1988), but twenty years later, as Foss and Klein’s (2008: 442) assessment concluded, none of these elements had really found their way into the contractual theory of the firm. The advancement of this agenda is not the object of this thesis, the aims of which are far more modest.

1.2 Aims and scope of the thesis

Although the aims of this thesis are not as ambitious as the proposal to integrate contractual and evolutionary perspectives on the firm, the implications may be at least as important. The argument is that by taking Coase’s first claim for granted and focusing exclusively on Coase’s second

7 Outside the contractual theory of the firm, these and other issues became central in the strategic management literature (e.g., Nelson, 1991; Conner, 1991; Barney, 1996; Grant, 1996; Madhok, 1996; Teece, Pisano and Shuen, 1997; Phelan and Lewin, 2000). See also the related literature on entrepreneurship (e.g., Casson, 2005; Langlois, 2007a; Foss and Klein, 2012).

8 See Foss (1993a) and Hodgson (1998a, 1998b) for a comparison of the contractual and evolutionary perspectives on the firm.
claim, contractual theorists have generally overlooked a key legal feature of the firm, without which theories of the firm are like Hamlet without the Prince. This vexing blind spot is revealed in Holmström’s (1999: 75) inability to provide a “well-developed explanation of asset ownership by firms” that would allow him to account for “one of the most significant and robust empirical regularities to be explained by any theory of the firm,” namely that “economic contracts are made with firms, not their employees or owners.” Holmström is justifiably puzzled because this significant and robust empirical regularity is at odds with the explicit or implicit assumption made in the theory of the firm literature, that the efficiency-enhancing central contractual agent is a flesh-and-blood human being (e.g., “entrepreneur,” “owner,” “employer”) or a group of such human beings (e.g., “owners,” “management”).

A case can be made for the fact that, along with other thorny questions that plague the literature, Holmström’s puzzle vanishes as soon as this erroneous conception regarding the identity of the efficiency-enhancing central contractual agent is discarded. The elementary institutional fact that most theorists of the firm have overlooked is that in order to become a fully operational firm in a modern market economy, an entrepreneur or an association of resource owners need to go through a formal registration or incorporation procedure. The outcome is this constitutive procedure is the

---

9 The expressions “registration” and “incorporation” are used synonymously throughout this thesis to designate the constitutive procedure by which the firm founders apply for and obtain the official recognition of the firm’s existence by the legal system. Although there are some variations both between common law and civil law jurisdictions, and within each kind of jurisdiction, these differences are minor. The principle that the firm
creation by the legal system of a distinct “legal person” in which ownership rights over assets used in production are vested, and in whose name contracts are made (Hansmann and Kraakman, 2000a: 390; Hodgson, 2002: 56; Gindis, 2007: 272; Pagano, 2010: 118; Deakin, 2012a: 115). With this assignment of legal personality, that Spulber (2009a: 152) has appositely described as a “foundational shift,” the legal system creates a “nexus for contracts” (Hansmann, 2013: 892, emphasis in original) and secures its continuity by “locking-in” the founders’ committed capital (Blair, 2004: 45), thereby allowing them to pledge assets, raise finance, and do business in the firm’s own name.

Given the basic principle to be found in any jurisdiction that only legal persons may own property and have the capacity to contract, and the implication that legally enforceable contracts can only exist between legal persons (Iacobucci and Triantis, 2007: 518; Robé, 2011: 3; Kornhauser and MacLeod, 2013: 918), it is something of a paradox that the notion of legal personality is absent from the prevailing narrative in the contractual theory of the firm. This is the “nexus paradox” identified in this thesis. In part, this state of affairs is imputable to the widespread reluctance among economists to rely on legal concepts of the firm, and to sustained efforts to define the firm with little or no reference to law. Little has changed since Masten (1988: only really comes into existence as a result of this constitutive procedure is the same in all jurisdictions.

10 The expressions “legal person” (or “juridical person”) and “legal entity” are used synonymously throughout this thesis, as are the expressions “legal personality” (or “juridical personality”) and “legal entity status.”
observed that “economists have either downplayed or rejected outright
the role of the law in defining the firm.” But the main reason for the
conspicuous absence of legal personality is the lasting influence of Jensen
and Meckling’s (1976) dismissal of the notion as an inconsequential “legal
fiction” in one of the most highly-cited articles published in economics since
1970.\textsuperscript{11}

Indeed, Jensen and Meckling’s (1976: 311) definition of the firm as a
“legal fiction which serves as a nexus for a set of contracting relationships
among individuals,” where by legal fiction is meant “the artificial construct
under the law which allows certain organizations to be treated as individuals”
(Jensen and Meckling, 1976: 310, n.12), is as celebrated as their immediate
rejection of the “personalization” of the firm. Their warning against the
anthropomorphic fallacies implied in discussions of the firm’s “objectives”
or “responsibilities” conveyed the idea that there is something “unnatural”
about the legal personality of firms. Although their ambiguous argument fails
by any standard of rigorous analysis, this interpretation seems to have stuck,
perhaps because Jensen and Meckling’s (1976: 311) view that “it makes little
or no sense to try to distinguish those things which are ‘inside’ the firm …
from those things that are ‘outside’ of it” was perceived to be an implication
of their definition of the firm. Given the literature’s focus on boundaries,
Jensen and Meckling’s article came to be much cited but little used, to
paraphrase Coase (1972: 63), except as a convenient foil, and the concept of

\textsuperscript{11} In addition to being the third most-cited article in economics (Kim, Morse and
Zingales, 2006: 192), Jensen and Meckling (1976) is the third most-cited article in finance
legal personality disappeared from the theoretical narrative as furtively as it had appeared.\textsuperscript{12}

The thesis shows that the fears distilled by Jensen and Meckling are unfounded once the meaning and functions of legal personality are properly understood. In order to disentangle the issues involved, it contextualizes Jensen and Meckling’s claims, revealing that their rhetoric was an anti-regulatory response to criticism of corporate America best represented by Nader, Green and Seligman’s (1976a) call for federal legislation to “tame the giant corporation.” Nader and his followers considered that key attributes of corporations, particularly legal personality, are privileges conceded by the state, and used this to justify further regulation. Jensen and Meckling (1976: 357) countered this reasoning by showing that, far from being state creations, private corporations, like other institutional arrangements, are the product of market forces, and took this to imply that regulations that increase costs for the human beings involved can only diminish the corporation’s chances of survival (Meckling and Jensen, 1977: 40; Jensen and Meckling, 1978: 32).

The explicit link between legal personality, misleading anthropomorphism and disputable political ends was picked up and amplified in the rising literature devoted to the economic analysis of corporate law (e.g., Hessen, 1979a; 1979b; Posner and Scott, 1980; Klein, 1980, 1982; Fischel, 1982a; Baysinger and Butler, 1985a; Butler, 1989; Easterbrook, 1989; Easterbrook and Fischel, 1989; 1991; Butler and Ribstein, 1995). In the process these ideas were accorded “the weight of scientific truth,” as

\textsuperscript{12} Until very recently no contractual theorist of the firm even mentioned the concept of legal personality. Today, exceptions include Spulber (2009a) and Van den Steen (2010).
Bratton (1989a: 409) observed, in what Kornhauser (1989: 1449) described as a “revolution [that] swept the legal theory of the corporation.” For years, as a result, anyone interested in both the theory of the firm and the topic of legal personality would have found little else on the subject in the law and economics literature. Yet none of these ideas were quite as scientific or revolutionary as it might have seemed. In fact, the 1970s and 1980s debate was a replica of an ancient dispute about the nature and origins of “corporate personality,” a dispute going back to Roman law that has been recurring ever since (Iwai, 1999: 584; Avi-Yonah, 2005: 771; Gindis, 2009: 26).

Notwithstanding important differences of context and emphasis, essentially the same debate structured discussions of both early corporate forms (monasteries, universities, charities, municipalities, guilds) and modern corporate forms (associations, trade unions, political parties, business corporations). The thesis surveys the history of this debate, that Iwai (2007: 243) referred to as “one of the most celebrated controversies in legal theory and legal philosophy,” demonstrating that theorists of the firm have much to learn from the “corporate personality controversy” of the past. In doing so it identifies the insights that can be used to dispel the numerous enduring

---

13 Within this literature, alongside Blair and Stout’s (1999) initial attempt to rehabilitate legal personality, legal entity status plays several important “functional” roles in the relatively recent work on corporate law and “organizational law” more generally developed by Hansmann and Kraakman (2000a, 2000b, 2004), and others. See Armour, Hansmann and Kraakman (2009) for a general overview.

14 The expression “corporate personality” normally refers to the legal personality of various corporate forms, and is sometimes misleadingly called “artificial personality” to mark the difference with the “natural personality” of human beings.
misconceptions and equivocations surrounding the notion of legal personality. In a nutshell, the aim of the thesis is to use these insights to propose a legally-grounded view of the nature and boundaries of the firm that fits the Coasean narrative while recognizing in law’s provision of legal entity status a fundamental institutional support for the firm. The aim, in short, is to have Hamlet with the Prince. The thesis is structured in the following manner.

1.3 Structure of the thesis

The thesis begins with a discussion of the theory of the firm literature. More precisely, Chapters 2 and 3 highlight the debates surrounding the boundaries of the firm that have resulted from the focus on Coase’s second claim, while using as a thread the issue of the identity of the central contractual agent. Contrary to other reviews, where the accent is mainly on different details in various models (e.g., Foss, Lando and Thomsen, 1999; Garrouste and Saussier, 2005; Gibbons, 2005), the thesis highlights the general thrust and overall continuity of the basic Coasean narrative.15 Chapter 2 outlines the

15 For the sake of overall coherence, some relatively minor contributions to the contractual theory of the firm, for example Baker, Gibbons and Murphy’s (1999, 2002) “implicit contracts” approach, have been excluded. Although it is a major contribution to the contractual paradigm, the so-called “positive agency theory” introduced by Jensen and Meckling (1976) and further developed by Fama and Jensen (see Fama, 1980; Fama and Jensen, 1983a, 1983b) is also excluded because, as Foss (2000: xxxvi) rightly observed, “principal-agent theories are not theories of the firm per se” and are not designed to answer
foundations of the field, constrating Coase’s (1937) original setup based on a sharp distinction between firms and markets (2.1) with Alchian and Demsetz’s (1972) influential objection that firms should be viewed as kinds of markets (2.2). It then addresses the problems raised by Richardson (1972) and others, according to which the traditional distinction between firms and markets may be too simplistic, as well as Cheung’s (1983) argument that the very idea of drawing the boundaries of the firm may be nonsensical (2.3). Overall, as will be clear from Chapter 2, a consensus on the nature and boundaries of the firm was not reached after this first wave of comparative institutional analysis.

The reactions to these problems are discussed in Chapter 3 that focuses on the three main efforts to reclaim firm boundaries. Based on Klein, Crawford and Alchian’s (1978) discussion of asset specificity, the new turn in the theory of the firm narrative owes much to Williamson’s (1979a) efforts to advance comparative institutional analysis by suggesting that the distribution of transactions between firms, markets, and other possibilities tends to match the characteristics of transactions, particularly the associated levels of relationship-specific investments (3.1). A similar theme was explored in Grossman and Hart’s (1986) formal model of firm boundaries based on the role played by nonhuman asset ownership (3.2). Concluding these efforts is Rajan and Zingales’s (1998) extension of the discussion to human capital-intensive firms in which many of the institutional arrangements that proved to be useful in physical capital-intensive firms may need to be revised (3.3). Curiously, it will be argued, this extension brings
the difficulties of defining the nature and boundaries of the firm back to the center stage.

In the following Chapters 4 and 5 the gap in the Coasean narrative is addressed by linking Coase’s first claim about the efficiency-enhancing properties of a central contractual agent and the issue of legal personality. Chapter 4 begins by assessing the numerous conceptual and definitional problems that afflict the theory of the firm literature (4.1). The difficulties involved in defining the boundaries of the firm are shown to be at least in part the result of the highly ambiguous uses of other key terms, such as “market” and “transaction costs,” and a preliminary case for using business registration to identify the firm is made. Chapter 4 then argues that Holmström’s (1999) puzzle and other unanswered questions regarding ownership, boundaries and the sources of intra-firm authority, can be resolved by recognizing that the central contractual agent is the legal person created for the firm founders by the legal system, rather than a human being or a group of human beings (4.2). The influence of Jensen and Meckling’s (1976) vigorous rejection of legal personality as leading to anthropomorphic fallacies and state regulation of private business, an accusation intended to discredit Nader and other social critics, is discussed next (4.3). This critique, as will become clear in the remainder of the thesis, is far from conclusive.

In Chapter 5 the dismissal of legal personality is put into perspective and related to the corporate personality controversy of the past. The chapter begins with a discussion of the old theory of corporations as state creations, precisely the conception that Nader attempted to resurrect, contrasting it with the approach introduced by Morawetz (1882) and others after the proliferation of general incorporation laws. On this view, that is strikingly
similar to Jensen and Meckling’s contractual approach, corporations are privately created associations of shareholders (5.1). Both approaches are then contrasted with Gierke’s (1900) theory of “real corporate personality,” that also emphasized the private origins of corporations while succumbing to anthropomorphic fallacies, but which was in turn purged from the latter claims by a new generation of British corporate theorists working around Maitland (5.2). Finally, the chapter considers parallel discussions in America, where Freund (1897) provided the earliest “rational study” of the value of corporate personality, to borrow Holmes’s (1897: 1001) expression (5.3). Overall, these considerations reveal that the notion of legal personality does not necessarily imply anthropomorphism and disputable calls for state control of private business.

Concluding the thesis, Chapter 6 uses these insights to make sense of recurring misunderstandings surrounding the notion of personhood, and proposes a legally-grounded view of the nature and boundaries of the firm that recognises the functional role and economic value of legal personality. When legal personality is understood as “the reference point for the enabling framework expressed economically as a nexus of contracts” (Schanze, 2006: 73) that literally carries the organizational structure of the firm (see Deakin, 2012a: 115; 2012b: 352) and allows it “to rank as a unit in the legal scheme,” as Carr (1905: 185) put it long ago, most if not all of the undesirable emotional equivocations can be discarded (6.1). A legally-grounded view of the kind suggested here not only allows us to have Hamlet with the Prince in order to further our understanding of the institutional structure of production (6.2) but also helps formulate future research agendas (6.3). To the extent that the argument of this thesis is successful, then all that it will have done, to paraphrase Coase (1992: 713), is “to urge the inclusion in our analysis of
features of the economic system so obvious that … they have tended to be overlooked.”
Chapter 2. From dichotomy to continuum

The issues surrounding what has been called “the firm” and the scope of economic behavior encompassed by Coase’s thesis are far from resolved.

Cheung (1983: 2)

Developed in Coase’s footsteps, the original contributions to the contractual theory of the firm discussed in this chapter revolve around two distinct views of firms and markets: the “dichotomy thesis” and the “continuum thesis.” According to the former, there are important differences between firms and markets, and firms are viewed as properly distinct from markets. In other words, intra-firm relations are seen as inherently different from inter-firm relations, implying that the mechanism coordinating individual behaviour within firms is quite different from the coordinating mechanism to be found outside the firm. Given that outside the firm the price system coordinates exchange, the theoretical problem is to demonstrate how an alternative mode of coordination replaces the price system within the firm. Having identified
these alternative modes, dichotomy theorists proceed to define the conditions under which firms are to be preferred to markets, or vice versa. At the margin, one is then able to draw the boundaries of the firm.

Hence Coase (1937) and other dichotomy theorists, most notably Williamson (1975), have focused on the relative advantages of firms and markets by comparing internal employment relations and external commercial contracts, or by considering the motivations behind the “make-or-buy” decision of vertical integration (2.1). This view of the boundaries of the firm is exactly what proponents of the continuum thesis deny based on several important arguments. One of these stems from Alchian and Demsetz’s (1972) claim that firms are essentially kinds of markets (2.2). The other, related, critique is that it may be difficult to identify the firm in non-arbitrary ways among the broad spectrum of contractual arrangements that individuals choose to adopt (2.3). What is at stake here is crucial, since the debate is really about the proper definition of the firm. Curiously, facing this difficult problem, a number of key theorists sidestep definitional matters altogether, and accept Cheung’s (1983) strong arguments against the use of the term “firm.”

### 2.1 Firms versus markets

Williamson’s (1975: 20) dictum, “in the beginning there were markets,” implies that firms and similar social institutions arise in a market economy in response to various shortcomings of the market allocation of scarce resources. Significantly, the “markets and hierarchies” research program
associated principally with Williamson (1971, 1973, 1975), and the contractual theory of the firm more generally, took shape in the context of the rising literature on “market failure” and “non-market allocation” (e.g., Kahn, 1966; Arrow, 1969; Radner, 1970; Akerlof, 1970; Spence and Zeckhauser, 1971; Hurwicz, 1973), a context in which Coase’s (1972: 73) agenda-setting proposal for research in industrial organization reminded economists that “market arrangements are the alternative to organization within the firm.” In fact, as Coase (1937) had famously argued in “The Nature of the Firm,” an article widely considered to be the first to propose the analytical framework that defines the field today, “the distinguishing mark of the firm is the supersession of the price mechanism” (Coase, 1937: 389).16

Contrary to his contemporaries, who focused on the cost and supply curves inherited from Marshallian microeconomics and on Robinson’s (1933) extension of the discussion to imperfect competition (see Williams, 1978; Moss, 1984), Coase’s (1937) aims were to account for the existence of firms in a market economy, to specify the determinants of the size of the firm, and to explain the distribution of activities between firms and markets. Developing his argument in three steps, Coase first introduced the previously unacknowledged disadvantage of the market, namely the “cost of using the price mechanism” (Coase, 1937: 390). He then explained the presence of the

---

16 This argument about the nature of the firm was only really rediscovered in the 1960s, after Coase’s (1960) discussion of social costs had made an impact (see Kitch, 1983: 202; Coase, 1988d: 33; Barzel and Kochin, 1992: 21). Nonetheless, at the dawn of the 1970s, Coase (1972: 63) could only observe that it was “much cited but little used.”
firm in a market economy by identifying the reasons for the firm’s cost advantage over the market. Finally, he explained the determinants of the firm’s size and the distribution of activities between the firm and the market by identifying the conditions that set the efficient boundary of the firm. The overall analysis, according to Coase (1937: 386-387), therefore became “tractable” by the most “powerful instrument of economic analysis,” namely, marginal substitution.\textsuperscript{17}

Coase’s major innovation was his claim that recourse to the price system, even in the most mundane of its forms, is never costless.\textsuperscript{18} “Marketing costs,” as Coase (1937: 392) called them, are necessarily incurred in all market transactions, and are pervasive at every stage of the contractual process since they include: the costs of discovering the relevant prices, the costs of negotiating, the costs of concluding separate contracts for each exchange transaction (see Coase, 1937: 390-391), as well as the costs of inspecting the execution of the contractual terms, as Coase (1960: 15) later added.\textsuperscript{19} Although this argument was born in the context of Coase’s discussion of the firm, its importance is not limited to the theory of the firm.

\begin{footnotesize}
\begin{enumerate}
\item With the exception of Rajan and Zingales’ approach discussed in Chapter 3, all contractual theories of the firm proceed in this manner.
\item A somewhat similar idea was introduced by Hicks (1935) in the field of monetary economics to explain the demand for money (see Klaes, 2000a: 193ff).
\item Based on Coase’s work, Dahlman (1979: 147-148) provided the classic typology of transaction costs as (a) search and information costs, (b) bargaining and decision costs, and (c) policing and enforcement costs, corresponding to successive stages of the contractual process.
\end{enumerate}
\end{footnotesize}
Indeed, the discussion of the effects of positive “transaction costs,” as they came to be called, is the hallmark of new institutional economics, and has fundamentally changed economic theory since the 1960s.\textsuperscript{20} Coase’s role in this genuine revolution within economic theory cannot be understated (see Kitch, 1983; Medema, 1995; Landes and Lahr-Pastor, 2011).

As Coase (1959: 27; 1960: 8) famously demonstrated, in the world of conventional price theory that assumes zero transaction costs, property rights, liability rules, and legal rules more generally are necessarily irrelevant, simply because resources tend to gravitate to their most valued uses through costless bargaining.\textsuperscript{21} In Coase’s (1960: 15) words, if “market transactions are costless … a rearrangement of rights will always take place if it would lead to an increase in the value of production.” Given this state of affairs, the distribution of economic activities between firms and markets is irrelevant and

\textsuperscript{20} Early uses of the notion of transaction costs include Malmgren (1961: 401), Sosnick (1961: 1351), Kessel and Alchian (1962: 522), Demsetz (1964: 12), Lees and Rice (1965: 143), Arrow (1965: 155), and Alchian (1965a: 820). By the 1970s, the notion had been introduced into general equilibrium models (e.g., Foley, 1970; Hahn, 1971). Interestingly, as Klaes (2000b: 569) has pointed out, Coase did not use the expression “transaction costs” before 1974 (see Coase, 1974a: 494).

\textsuperscript{21} Stigler (1966: 113) famously called this result the “Coase theorem”, although his formulation was different (see Medema, 1999: 213). One of the clearest formulations is arguably Calabresi’s (1968: 68, emphasis in original): “if people are rational, bargains are costless, and there are no legal impediments to bargains, transactions will ex hypothesis occur to the point where bargains can no longer improve the situation.” The importance of this argument about costly exchange is that it reveals the “missing chapter” of pre-Coasean exchange theory (Hirshleifer, 1973). Detailed discussions of the Coase theorem can be found in Farber (1997), and Medema and Zerbe (1999).
unimportant. On the other hand, once it is realized that various market operations are often extremely costly, sufficiently costly at any rate to prevent many transactions that would be carried out in a world in which the pricing system worked without cost” (Coase, 1960: 15), it becomes clear that property rights, liability rules, and the distribution of activities between firms and markets, can affect efficiency and are thus highly relevant to the economist’s inquiry.

On several occasions throughout his long career, Coase explained what he saw as the most fundamental flaw of conventional economic theory, that he disapprovingly labelled “blackboard economics” (Coase, 1964: 194; 1988c: 28ff; 1992: 714). Blackboard economics, with its excessive commitment to the frictionless or zero transaction cost world of competitive price theory, Coase (1960: 43) argued, has prevented economists from comparing real institutional alternatives, all of which contain imperfections. As Demsetz (1969a: 1-2, 19-20) put it, picking up on Coase’s idea, when economists compare the ideal, perfect, costless market to some empirically observed institutional arrangement they are committing a “nirvana fallacy,” that ultimately leads them to misguided policy prescriptions. On the contrary, Demsetz argued, the proper economic analysis of institutions developed in Coase’s footsteps needs to adopt the “comparative institutional approach” where imperfect institutional arrangements are compared with one another. In

---

22 If markets lead to a perfectly efficient allocation of resources, the existence of firms and other institutions is “redundant, or worse, inefficient” (McManus, 1972: 37).
general, this methodological strategy is characteristic of the new institutional economics (see Williamson, 1979a: 261).\textsuperscript{23}

The second part of Coase’s theory of the nature of the firm is set in this new framework of a positive transaction costs world – the only world economists should be concerned with, since it is the only world that actually exists – where it becomes possible to identify the transaction cost-economizing role of the firm. After all, if firms, along with other institutions, were unable to reduce costs compared to the market, there would be no economic justification for their existence. Given Coase’s (1937: 389) claim that within the firm the price mechanism is superseded, the theoretical challenge he faced was to explain not only how individual behavior is coordinated in the absence of the coordinating role usually assumed by the price system, but also how this alternative mode of coordination allows the reduction of transaction costs. This latter question is fundamental for all the theorists of the firm discussed hereafter, even for those who deny Coase’s dichotomy thesis.

For Coase (1937: 388), quoting Robertson’s (1923: 85) colorful phrase, a theory of the firm needs to explain why “we find ‘islands of conscious power in th[e] ocean of unconscious co-operation like lumps of butter coagulating in a pail of buttermilk’.” In terms that Coase himself did not use, why do we observe the “visible hand” of management given the efficiency-

\textsuperscript{23} Coase argued that blackboard economics is “pointless” (Coase, 1984: 230), even when it is carried out with “great ingenuity” (Coase, 1988c: 28), because the world it depicts “lives in the minds of economists but not on earth” (Coase, 1992: 714). See Mäki (1998) for a detailed discussion of Coase’s criticism of unrealistic economic theory.
enhancing power of the “invisible hand” of markets? Like most economists of his day, who were discussing the nature of the entrepreneurial function (e.g., Knight, 1921; Robinson, 1934; Kaldor, 1934), Coase (1937: 393) defined the firm as a “system of relationships that comes into existence when the direction of resources is dependent on an entrepreneur.” The novelty of Coase’s answer was his suggestion that when production is carried out “by forming an organisation and allowing some authority (an ‘entrepreneur’) to direct resources, certain marketing costs are saved” (Coase, 1937: 392). Firms, Coase argued, are viable because they reduce the number of transactions necessary for the organization of production, thereby reducing the associated transaction costs. Coase (1937: 391) explained:

Contracts are not eliminated when there is a firm but they are greatly reduced. A factor of production (or the owner thereof) does not have to make a series of

---

24 Coase’s question needs to be set against the background of the so-called “socialist calculation” debate of the 1920s and 1930s that revolved around the relative efficiency of centralized planning as opposed to decentralized markets. Citing Hayek (1933) on these two kinds of order, Coase (1937: 387) observed that “in searching for a definition of a firm” the question that needs to be addressed is whether “economic planning” can improve on “what is already done by the price mechanism.” See Coase (1988b: 8; 1992: 715) on how the debates spawned by the Russian Revolution influenced his thinking. Interestingly, Cheung (1982a:39) later pointed out that some “sophisticated observers” of China have defended its centralized system “in almost exactly the same terms that Coase and his followers use to explain the firm.”

25 Knight (1921: 268) called “cephalization” this centralization of decision-making and control.
contracts with the factors with whom he is co-operating within the firm, as would be necessary, of course, if this co-operation were as a direct result of the working of the price mechanism. For this series of contracts is substituted one.\textsuperscript{26}

In more formal terms, once activities are centralized under the administrative authority of an entrepreneur, it is easy to see that, above a certain number of factor owners, the number of contracts needed to organize the production process is reduced. Indeed, if every one of \( n \) independent resource owners has to contract with every other owner \((n - 1)\), which would be case in the open market, the total number of commercial contracts is \( n(n - 1)/2 \). Clearly, when an entrepreneur acts as the sole central contracting party, assuming that the entrepreneur is one of the \( n \) owners, only \( n - 1 \) individual contracts linking each remaining owner to the entrepreneur are necessary. For any \( n \geq 3 \), this number is strictly less than \( n(n - 1)/2 \). The greater the number of owners involved the greater the reduction in necessary contracts.\textsuperscript{27} As Coase (1988d: 39) later observed, the emergence of the firm leads to “much less complicated contractual arrangements.”\textsuperscript{28} By reducing the associated

\begin{itemize}
\item \textsuperscript{26} As Coase (1960: 16) clarified: “within the firm individual bargains between the various cooperating factors of production are eliminated and for a market transaction is substituted an administrative decision. The rearrangement of production then takes place without the need for bargains between the owners of the factors of production.”
\item \textsuperscript{27} Related formulations can be found in Williamson (1975: 46), Milgrom and Roberts (1992: 331), and Spulber (1999: 263-264).
\item \textsuperscript{28} Arrow’s (1974: 68ff) discussion of information and communication costs in
\end{itemize}
transaction costs, such an arrangement is unambiguously efficiency-enhancing.

On Coase’s view, efficiency is also enhanced because the nature of the unique contract that ties each resource owner to the entrepreneur is quite distinct from the typical commercial or sales contract to be found in ordinary market exchange. Specifically, Coase argued, under the terms of this contract the owner of a factor of production relinquishes some autonomy and accepts the entrepreneur’s authority, “within certain limits,” on a long-term basis in exchange for some defined remuneration (Coase, 1937: 391). Given their unspecified duration, he continued, such contracts “only state the limits to the powers of the entrepreneur” (Coase, 1937: 391), and are “expressed in general terms, the exact details left until a later date” (Coase, 1937: 392). Accordingly, Coase (1937: 403-404) concluded, since these contracts bear a strong resemblance to the “legal concept of ‘employer and employee’,” the suggested theory of the firm is “realistic in that it corresponds to what is meant by the firm in the real world” (Coase, 1937: 386).

organizations likewise underlined the efficiency-enhancing properties of centralization. This line of comparative institutional analysis was pursued by Bolton and Dewatripont (1994), Laffont and Martimort (1997), and Dessein (2002).

29 In modern language, one would say that these contracts are “incomplete.”

30 Following Coase, many economists have argued that employment relations, where subordinates obey their superiors’ orders, are essential to the definition of the firm. Inevitably, this raises the “Hobbesian question” of the theory of the firm: why would a free and sovereign individual, an owner of a valuable resource, accept the authority of some other agent, even if the limits of that authority are stipulated by contract? Coase was somewhat vague. A standard economic answer is that, to the extent that one observes
Interestingly, Coase (1988c: 37) later saw this insistence on long-term employment contracts as the main weakness of his 1937 article. Instead, he formulated another way of looking at the economizing role of the firm. The emergence of the firm, he observed, leads to “the substitution of interfirm for factor-factor transactions and of firm-consumer for factor-consumer transactions” (Coase, 1988c: 39). To see why firm-consumer transactions reduce costs in comparison to factor-consumer transactions, one needs only to think of the number of contracts necessary for an exchange to occur when a consumer needs to negotiate with a large number of stand-alone factors (i.e., specialized agents producing for others) for each component and their assemblage, and to compare this with a situation in which the consumer can contract with a representative of a coalition of resource owners, a central agent, for the complete commodity. Overall, for Coase, the emergence of a central contracting party simplifies not only internal relations between resource owners (Coase, 1937) but also external relations between third parties and the group of resource owners as a whole (Coase, 1988c).

superior-subordinate relations, (a) the interests of each subordinate must be served at least as well as elsewhere, and (b) there must be some “area of acceptance” (Simon, 1951: 294) or “zone of indifference” (Barnard, 1938: 167ff) within which subordinates indifferently carry out their superiors’ orders. It is also clear that (c) the superior’s authority will tend to be exercised within that zone given the subordinate’s exit rights. For related discussions of the employment relation, see March and Simon (1958: 90ff), Williamson, Wachter and Harris (1975: 268ff), and Williamson (1975: 71ff; 1985a: 218-219). A more general overview of the notion of “Hobbesian hierarchy” in economics and political science is provided by Lake (2009).

31 This point is discussed in more detail by Eggertsson (1990: 162-163).
Coase presented the third and final part of his argument in the form of a question: if the firm has these efficiency-enhancing properties, why is not all production carried out within one big firm (Coase, 1937: 394)? In other words, why are there markets?\textsuperscript{32} The firm’s size will tend to grow, Coase reasoned, as long as it is worthwhile to internalize a transaction previously organized by two or more entrepreneurs, that is, as long as the benefits of vertical integration outweigh the costs of the corresponding market transactions. Inevitably, however, as additional transactions are brought within the scope of the entrepreneur’s authority the costs of internal organization will tend to rise as a function of the size of the firm (Coase, 1937: 395),\textsuperscript{33} and this is what will ultimately determine the efficient boundaries of the firm: boundaries will tend to be set at those points at which the costs of using the price system are equal to the cost of internal organization.\textsuperscript{34} Accordingly, Coase (1937: 404) concluded, “the principle of

\textsuperscript{32} Williamson (1985a: 132-133) called this a “chronic puzzle,” while Stiglitz (1991: 18) dubbed this the “centralization paradox.”

\textsuperscript{33} Coase (1937: 395-396) accepted the then standard neoclassical explanation that the size of the firm is limited by the “diminishing returns to management” (e.g., by the tendency of the entrepreneur to make mistakes as the number of transaction to be organized increases). Put differently, administrative or bureaucratic costs limit the size of the firm. However, contrary to his contemporaries who ignored transaction costs, Coase’s (1937: 403) account of the boundaries of the firm demonstrated how a theory of the firm could address the relative costs and benefits of both firms and markets.

\textsuperscript{34} Coase (1972: 64) argued that the organization of industry is determined not just by technological factors but also by considerations of this kind.
marginalism works smoothly,” and the analysis is not only realistic but also “tractable.”

Coase’s followers extended and refined the analysis. The most important subsequent writer working within Coase’s dichotomy framework is unquestionably Williamson, whose early discussions of vertical integration or the “make-or-buy” decision (Williamson, 1971: 114; 1973: 316) and the employment relationship (Williamson, Wachter and Harris, 1975: 260) were explicitly couched in a comparative institutional framework. Williamson’s (1975) *Markets and Hierarchies* included applications of the basic framework based on the twin notions of “market failure” and “organizational failure” to a new set of questions, including antitrust, market structure, organizational innovation, and much more.35 Notwithstanding important differences with both Coase and Williamson, Arrow’s (1974) *Limits of Organization*, that presents firms as “means of achieving the benefits of collective action in situations in which the price system fails” (Arrow, 1974: 33), was also written in the spirit of the dichotomy thesis.

Instead of discussing Williamson’s contributions to the dichotomy paradigm, as would seem natural at this point, a rarely noticed aspect of Coase’s theory needs to be brought to the center stage. As we have seen, the most crucial aspect of the efficiency-enhancing structure of the firm is that contractual relations between resource owners are not really contractual relations between resource owners. Instead of having to contract with each other, resource owners “joining” the firm need only to contract with the

35 See Williamson and Ouchi (1981) on the origins, implications and prospects of the research programme.
entrepreneur. At no point, however, does Coase specify the identity of the agent, among the $n$ resource owners, who is to take on the role of the entrepreneur, thus acting as the central contracting party that coordinates production. Indeed, Coase does not even say if the entrepreneur is one of the $n$ original resource owners.\(^{36}\) Coase simply says that, for understandable efficiency reasons, one agent needs to be the central contract party, and that agent can be called the “entrepreneur.”

Generally speaking, for the Coasean setup to work, it does not really matter whether it is the provider of capital or the provider of labor that acts as the entrepreneur. This leads Cheung (1987: 56) to observe that “it is not clear whether it is the entrepreneur who employs the workers or the workers who employ the entrepreneur.”\(^{37}\) It follows that Coase’s setup is consistent with the standard neoclassical logic famously captured by Samuelson (1957: 894), that in competitive markets “it really doesn’t matter who hires whom: so have labor hire ‘capital’.” As is well known, this observation, among others things, spawned considerable reactions from neo-Marxist or “radical” political economists in the late 1960s and early 1970s. According to economists belonging to what Lindbeck (1971: 2) called the “New Left movement,” the Smithian account of the efficient division of labor implicit in Samuelson’s statement is historically absurd. Indeed, they argued, capital

---

\(^{36}\) If the entrepreneur is not one of the original $n$ resource owners then $n(n - 1)/2$ is strictly greater $n$ individual contracts for every $n > 3$.

\(^{37}\) As Ellerman (2007: 16) commented, “anyone, by becoming the hiring party, the nexus of the hiring contracts, can become the firm.”
hires labor for reasons that have less to do with efficiency than with the capitalists’ thirst for power and control.

Expressing this “radical challenge” to orthodoxy (see Franklin and Tabb, 1974; Cain, 1974, 1975), Marglin (1974) re-examined the move from the putting-out system to the factory system during the nineteenth century, demonstrating that centralized organization of production has always been designed to ensure the entrepreneur’s essential role in the production process, thereby guaranteeing him a share of the pie. Moreover, since the centralization of workers in large factories allowed greater control of labor by facilitating supervision and punishment, the purpose of the factory system was to increase the entrepreneur’s share of the pie (see Marglin, 1974: 95).38 Although, as Lindbeck (1971: 1, emphasis in original) pointed out, “it is rather improper to refer to the economics of the New Left,” since “this movement has no well-defined and unified economic policy program,” radicals did generally converge on the defense of more “socialist” forms of organization of the production process in which labor hires capital.39 Like

38 For the radicals, the process of “depriving the workers of control of product and process” (Marglin, 1974: 62) had already begun with the passage from the completely decentralized craft production system that prevailed before the industrial revolution to the putting-out system, a decentralized organization of domestic workshop production that was centrally controlled by a capitalist-merchant who had interposed himself between the workers and the market, thereby pocketing the proceeds. The passage from the putting-out system to the factory system, centrally controlled by a new breed of capitalist-industrialists, was merely the more advanced stage of this historical process.

39 The New Left literature includes Edwards et al. (1970), Edwards, Reich and Weisskopf (1972), Sherman (1972), Braverman (1974), Bowles and Gintis (1975), Gintis
Ward (1967), Vanek (1970) and other proponents of “market socialism,” radicals viewed worker cooperatives and other “democratic” decision-making schemes as viable alternatives to “capitalist” firms.

An important part of the identity of the contractual theory of the firm comes from this debate with the radicals (see Putterman, 1986: 25). Prominent new institutionalists, including Furubotn and Pejovich (1972), Williamson (1975), and Jensen and Meckling (1979), responded that hierarchy and more generally “capitalist” firms, in which capital hires labor and not the other way around, are relatively more efficient. Adopting a comparative institutional approach, Pejovich (1976), Williamson (1980), De Alessi (1980), Furubotn (1981), and many others, argued that labor-managed firms are plagued by the so-called “tragedy of the commons,” to use Hardin’s (1968) celebrated expression: where property is owned in common by workers, each worker has the incentive to “shirk” or to “free-ride,” since a rational individual can reduce effort without a proportional income reduction (see Pejovich, 1990: 193ff; Eggertsson, 1990: 165ff; Milgrom and Roberts, 1992: 294ff). For new institutionalists, externalities of this kind are greatly reduced by the institution of private property rights.

---


40 Detailed historical rebuttals of the radicals’ account of the Industrial Revolution were also provided by members of the so-called “New Economic History” movement, such as Clark (1984) and Landes (1986).

2.2 Firms as markets

In “Production, Information Costs and Economic Organization,” written in the midst of the debate with the radicals, Alchian and Demsetz (1972: 784) pointed out that Coase’s (1937) theory of the firm was silent on ownership issues, and that this made it difficult to apply it in a straightforward manner to the capitalist firm. Accordingly, Alchian and Demsetz’s aim was to update and focus the theory.\(^{42}\) In the process, leading to what is arguably the second most important contribution to the contractual theory of the firm, Alchian and Demsetz (1972) explicitly denied nearly all of the essential features of the firm highlighted by Coase. Overturning Coase’s main claim, Alchian and Demsetz (1972: 795) presented “the firm and the ordinary market as competing types of markets,” and argued that market transactions were not eliminated within the firm. This objection to the dichotomy thesis has had a lasting influence on the theory of the firm.\(^{43}\)

\(^{42}\) In this sense Alchian and Demsetz’s article can be seen as an implicit response to the radical challenge along the lines of Lindbeck’s (1972: 682) call to “prove to the New Left … that traditional economics is a powerful tool for understanding fundamental problems” related to ownership and power. Interestingly, both Arrow (1974: 15-16) and Williamson (1975: 45) explicitly contextualized their discussions of the efficiency of hierarchy as a response to the New Left’s critique of authority.

\(^{43}\) In addition to its twelfth position in Kim, Morse and Zingales’s (2006) study of what has mattered in economics since 1970, Alchian and Demsetz’s (1972) paper is listed in the “Top 20 Articles” published in the *American Economic Review* over the first century of its existence (Arrow et al., 2011).
In a heavily-cited passage, Alchian and Demsetz (1972: 777) observed 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.” In fact, they explained, an “employer” has no more “authority” over an “employee” than a customer has over a grocer. Given that “the employee can terminate the contract as readily as can the employer,” they continued, “long-term contracts … are not an essential attribute of the firm” (Alchian and Demsetz, 1972: 783). On the contrary, “employers” and “employees” are constantly renegotiating the terms of their interaction (Alchian and Demsetz, 1972: 777). In all cases worth considering agents are free to take their business elsewhere, and the power to withhold future business, whether exercised by an unsatisfied customer or an unsatisfied employee, is the only meaningful type of “power” or “punishment.”

Nevertheless, Alchian and Demsetz (1972: 783-784) argued that their view of the firm is “not necessarily inconsistent with Coase’s,” and proposed to “[take] a step down the path pointed out by Coase.” The main problem with Coase, they pointed out, is that his focus on transaction costs led him to ignore the fact that other types of costs may account for the emergence of the firm. Importantly, economizing on the internal “costs of managing” may confer an advantage to the firm at given levels of market transaction costs (Alchian and Demsetz, 1972: 783), indicating that the comparative

---

44 Of course, this argument assumes competitive markets where buyers and sellers (be it of labor services or of bread) have numerous trading alternatives. The argument also assumes that buyers and sellers are interchangeable, that is, that there is nothing specific about particular trading partners that may change the nature of their relations, alter their alternatives, and hence affect their incentives.
institutional analysis of firms and markets needs to include both external market transaction costs and internal management costs.\textsuperscript{45} The upshot is that taking these costs into account helps explain why firms come in a variety of forms of organization (e.g., corporations, partnerships, profit-sharing firms, socialist firms), and how the reduction in transaction costs can lead to the expansion of the firm, as evidenced by the move from the putting-out system to the factory system (Alchian and Demsetz, 1972: 784). Coase’s theory cannot accomplish any of these tasks.

For all their dissatisfaction with Coase, the key element of Coase’s theory of the nature of the firm was retained and given new meaning by Alchian and Demsetz. This is the idea of a central contractual agent. The significant difference between intra-firm relationships (a grocer and his employee) and ordinary market contracting (a grocer and his customer), they hypothesized, lies in a “team use of inputs and a centralized position of some party in the contractual arrangements of all other inputs” (Alchian and Demsetz, 1972: 778, emphasis in original). The theoretical challenge is 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” (Alchian and Demsetz, 1972: 794).

Generally speaking, as Alchian and Demsetz explained, for owners of inputs or factors of production to choose to join forces the joint output, or the

\textsuperscript{45} Demsetz (1988a: 146-147; 1997: 426; 2008: 107) repeatedly observed that it is important to distinguish the costs of using the price system from the costs of using the firm’s internal management system.
gain from cooperative behavior, needs to be greater than the sum of the outputs that would have otherwise been obtained. A “team” is a special case of cooperative behavior producing such a superadditive output, but in which individual contributions to the joint output cannot be easily detected. Indeed, contrary to the team’s observable joint output sold in the market, team member effort is private information that may be very costly to detect. It follows that each input owner has an “incentive to shirk” (Alchian and Demsetz, 1972: 780): free-riding, or under-investment in effort, may very well arise because agents realize that their effort can be reduced without a proportional income loss. This difficulty to establish a correlation between efforts and rewards characteristic of team production situations contrasts with markets that more often than not allow, according to Alchian and Demsetz (1972: 778), a “high correlation between rewards and productivity.”

It is important to notice that the problem described by Alchian and Demsetz is a variation on the tragedy of the commons: team members are tempted to reduce their individual effort in order to maximize their share of a

46 In Alchian and Demsett’s (1972: 779) words, “the production function is not separable.”

47 Alchian and Demsett (1972: 780) pointed out that “if detecting such behavior were costless, neither party would have an incentive to shirk.”

48 As Demsett (1967: 356) had argued, a great part of the success of markets lies in the fact that they effectively concentrate costs and benefits on the relevant resource owners, meaning that incentives to use resources efficiently are stronger in markets. This view of markets was developed in Coase’s (1960) footsteps by Alchian (1965a, 1967), Demsett (1964, 1966), and others. See also Alchian’s (1969a) discussion of resource unemployment resulting from information costs.
given, common pie. Of course, like in the tragedy of the commons situation, if all team members reduce effort, the pie will tend to disappear altogether, as will the rationale for the team’s existence. In such circumstances, a “multitude of bilateral exchange of separable individual outputs” (Alchian and Demsetz, 1972: 780) would be preferred to team production. Accordingly, viable teams will tend to be formed only when the possibility of shirking is somehow reduced. It follows that team production situations justify some form of management, “viewed basically as one of organization among people” (Alchian and Demsetz, 1972: 779), and that this is the crucial source of management costs ignored by Coase. From this point of view, it is not so much the “the costs of spontaneously negotiating contracts in the markets” (Alchian and Demsetz, 1972: 782) but “the costs of metering or ascertaining the marginal products of the team’s members [that calls] forth new organizations and procedures” (Alchian and Demsetz 1972: 780).49

For Alchian and Demsetz, in the presence of positive “measurement costs,” as they came to be called (e.g., McManus, 1975: 340; Klein, 1980: 362; Barzel, 1982: 48), it makes economic sense for “someone to specialize as a monitor to check the input performance of team members” (Alchian and Demsetz, 1972: 781), provided that the monitor is given the “incentive not to shirk as a monitor” (Alchian and Demsetz, 1972: 782), and that the joint output value thus produced is large enough to cover the costs of managing

49 Alchian’s (1965b: 34) earlier discussion of the managerial theory of the firm associated with Marris (1964) and Williamson (1964) had already emphasized the importance of the “costs to the stockholders of detecting and policing the manager’s behaviour and effectiveness.”
the team. From this perspective, they argued, the obvious efficiency-enhancing contractual arrangement between team members 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 (Alchian and Demsetz, 1972: 782).

The monitor’s disciplining role will be even more effective, they continued, if the monitor-residual claimant is granted the right to “unilaterally terminate the membership of any other member without terminating the team itself or his association with the team” (Alchian and Demsetz, 1972: 782-783). Clearly, a contractual arrangement of this kind implies that the monitor-residual claimant is “the central party common to all contracts” (Alchian and Demsetz, 1972: 783). The identity of the central contracting agent is further revealed by adding the right to change team membership, the right to expand or reduce team size, and the right to sell this entire bundle of rights. As Alchian and Demsetz (1972: 783, emphasis in original) put it: “it is this entire bundle of rights … that defines the

50 The “moral hazard” explanation of economic organization developed by Holmström (1979, 1982) builds on Alchian and Demsetz, but can be traced back to Knight (1921). Knight’s views were considered but set aside by Coase (1937: 394), who believed that Knight thought it “impossible to treat scientifically the determinants of the size of the firm,” precisely what Coase went on to attempt. Discussions of Knight’s theory of the firm can be found in Barzel (1987a), Demsetz (1988b), Langlois and Cosgel (1993), and Foss (1993b). A comparison between Coase and Knight is provided in Boudreaux and Holcombe (1989), and Demsetz (1995).

51 Residual income helps identify the entrepreneur for Knight (1921: 271ff). On Barzel’s (1987b) interpretation, residual income is the “entrepreneur’s reward for self-policing.” See also Demsetz (1988b).
ownership (or the employer) of the classical (capitalist, free-enterprise) firm.” The firm’s “owner” is the agent in whose hands has been assembled a specific bundle of rights, among which the roles of central contracting party and residual claimant.

Arguably, for this coalescence of rights to arise, team members need to realize that it is in their interest to relinquish some of their claims on the joint output value. In fact, the contractual arrangement is precisely the transfer of some of the property rights composing each member’s bundle of rights to the central agent known as the “owner.” To the extent that one should only observe such a transfer of rights when the expected benefits outweigh the costs, the additional value resulting from the definition of a new bundle of rights must be greater than both the management costs created by the shirking problem and the costs of specifying the new property rights structure. As Demsetz (1967: 350) had famously explained, new “property rights develop to internalize externalities when the gains of internalization become larger than the costs of internalization.”

---

52 This line of reasoning is a consequence of Alchian’s (1965a: 818ff) and Demsetz’s (1967: 347) earlier work on property as a partitionable bundle of rights that may be temporarily or permanently assigned to other persons. Cheung (1968, 1969, 1970) is another important contributor to this tradition building on Coase (1960). A concise overview of the property rights paradigm is provided by Alchian and Demsetz (1973).

53 Alchian (1969b: 353) similarly explained: “if the value of some right … rises … the costs of specification, identification and assignment of rights become more worth incurring.” Kleinsorge called this insight “Demsetz’s thesis” (Ashley, Kleinsorge and Kunreuther, 1967: 377), and it has since become known as the “Demsetz thesis.” Although it is generally attributed to Demsetz (1967), the thesis also appeared in Demsetz (1964: 19).
assignment of the rights and liabilities associated with private property systems overcomes the tragedy of the commons situation, then Alchian and Demsetz’s (1972) explanation of the rise of the capitalist firm, with its specific rights structure designed to discipline team members and to realize the benefits of team production, is a case in point.

The view that Alchian and Demsetz ultimately hold is that the capitalist firm is the long-run outcome of market forces. Given Alchian’s (1950) important evolutionary argument, the owner’s right to unilaterally alter the membership of the team without affecting the other contracts is fundamental since it allows the owner to progressively compose efficient teams in a trial-and-error process (see Alchian, 1950: 219). As Alchian (1984: 35) later explained, forming teams is “a process of sampling among potential members of a team, the contractual relations, and the environment.” The design of institutional arrangements, as Demsetz (1969b: 7) put it, should “encourage experimentation without overly insulating these experiments from the ultimate test of survival” in the marketplace (see also Demsetz, 1969a: 20). Since transferability or saleability of rights is a vital part of the definition of markets, and therefore an additional incentive to manage resources efficiently (see Demsetz, 1966: 62; Alchian, 1967: 375), the right to sell the entire bundle of rights is of utmost importance because it creates the conditions for a “market for firms.”

See Merrill (2002), and especially Lueck and Miceli (2007: 209ff). Interestingly, Demsetz (2011a, 2011b) has recently raised doubts about the validity of this explanation.

54 In Jensen and Meckling’s (1992: 257) words, the rights assignment problem is solved by alienability in markets.
The presence of information asymmetries, however, may reduce the effectiveness of the disciplining power this market, since potential buyers of the bundle of rights associated with the firm’s ownership compete in markets that may not adequately price the firm’s value.\textsuperscript{55} Alchian’s (1967: 374) view, that “insiders may know more than outsiders,” is in fact exactly what Alchian and Demsetz’s account of input monitoring had explained. To the extent that it is costly to measure insider performance, they explained, employees can act inefficiently, and this is why the emergence of a monitor specialized in detecting deviant behavior enhances efficiency. By definition, then, the monitor “acquires special superior information about th[e] productive talents” of inputs, and uses this additional knowledge as a “basis for superior decisions about efficient or profitable combinations of those heterogeneous resources” (Alchian and Demsetz, 1972: 793). Arguably, this is the main source of team productivity. Both Demsetz (1988a: 152; 1995: 17) and Alchian (1991: 233) later observed that the failure to discuss in more detail the source of team productivity was the main weakness of their 1972 paper.

Instead, Alchian and Demsetz (1972) concluded from the preceding that contractual relations within the firm do not differ “in the slightest degree from ordinary market contracting between any two people.” There is simply a sales contract. The employer “sells” the superior information obtained from monitoring employee-inputs “as he aids them in ascertaining good input combinations for team activity” (Alchian and Demsetz, 1972: 793), and “the

\textsuperscript{55} Nonetheless, Alchian and Demsetz (1972: 788) accepted Manne’s (1965, 1967) analysis of the disciplining power of the “market for corporate control” that acts as an additional constraint on managerial shirking (see also Alchian, 1969b: 344ff).
employee ‘orders’ the owner of the team to pay him money in the same sense that the employer directs the team member to perform certain acts” (Alchian and Demsetz, 1972: 783). The firm, Alchian and Demsetz (1972: 793) inferred, should be properly viewed as a “highly specialized surrogate market,” that is, as a “specialized market institution for collecting, collating, and selling input information.” Alchian and Demsetz (1972: 794-795) concluded with a “highly conjectural but possibly significant interpretation”:

As a consequence of the flow of information to the central party (employer), the firm takes on the characteristic of an efficient market in that information about the productive characteristics of a large set of specific inputs is now more cheaply available … Inputs compete with each other within and via a firm rather than solely across markets as conventionally conceived. Emphasis on interfirm competition obscures intra-firm competition among inputs … The firm is a device for enhancing competition among sets of inputs resources as well as a device for more efficiently rewarding the inputs. In contrast to markets and cities which can be viewed as publicly or nonowned market places, the firm can be considered a privately owned market; if so, we could consider the firm and the ordinary market as competing types of markets, competition between
private property markets and public or communal markets.\textsuperscript{56}

In one short article, Alchian and Demsetz defined the essence of the contractual theory of the firm. In the process, they provided a theoretical framework that could be used to defend the so-called classical capitalist firm. In a nutshell, the centralized contractual structure constitutive of ownership is said to be efficient, and its relative efficiency, compared with other types of organizational forms, is the result of competitive market forces.\textsuperscript{57} Externalities are best internalized by markets, and more precisely by the specification and assignment of rights. New specifications and assignments are literally extensions of the market. Given that within the firm externalities are also internalized by the specification and assignment of rights, firms are a special type of market. They are “market-like.” From this perspective, as

\textsuperscript{56} Defining firms as surrogate markets, Alchian and Demsetz explicitly challenged Coase’s dichotomy thesis but also tacitly suggested a contractual interpretation of Doeringer and Piore’s (1971) notions of “internal labor markets” developed within the “market segmentation” approach promoted by the New Left (see Reich, Gordon and Edwards, 1973). Contrary to Doeringer and Piore’s (1971: 1) definition of an internal labor market as “an administrative unit within which the pricing and allocation of labor is governed by a set of administrative rules and procedures,” Alchian and Demsetz discounted administrative rules and emphasized internal competition within firms instead.

\textsuperscript{57} As Holmström and Tirole (1989: 67) observed, “the limited extent of partnerships and cooperatives in our economy lends some support to the owner-monitor model, since free-riding could be a big problem in these organizations.” From this point of view, Holmström (1982: 325, 327) suggested that the Alchian and Demsetz model is about how efficiency is restored when the shirking-ridden labor-managed firm (but also the partnership) changes into a capitalist firm.
Alchian (2006 [1978]: 643) later clarified, “a firm is not an alternative to the market … Rather it is a result of contractual activity in the market.”

Contrary to appearances, the question of the identity of the central contractual agent is not fully resolved in Alchian and Demsetz’s story. Like Coase, they simply argued that, for understandable efficiency reasons, one agent needs to specialize in monitoring and assume the rights of ownership, thereby becoming the employer. Presumably, one of the original team members assumes these roles, but this does not really matter for their account. In the words of Alchian and Allen (1977: 206, emphasis in original), “there is no difference … between fishermen renting the boat or the boat owner hiring fishermen as employees.” This led Holmström and Tirole (1989: 69) to point out that “nothing would preclude the monitor from being an employee of a separate firm with a service contract that specifies his reward as the residual output. Similarly with workers. They could be monitored and paid as independent agents rather than employees.” Holmström and Tirole refer to this apparent irrelevance of institutional or legal setting for contracts as the problem of “organizational anonymity.”

58 Nevertheless, Alchian and Demsetz (1972) did not conclude that it is impossible to set the efficient boundary of the firm, that is, to make their analysis “tractable” in Coase’s sense. In a perfectly Coasean spirit, Alchian and Demsetz (1972: 794, n.18) observed, albeit in a footnote: “presumably, at some size of the firm, specialized knowledge about inputs becomes as expensive to transmit across divisions of the firms as it does across markets to other firms.” This argument was later developed by Demsetz (1995: 34ff).

59 This is also the case in Cheung’s (1983: 8) example of Chinese riverboat-pulling teams that hire monitors to wield the whip: the monitor may be any one of the team members, who possibly take turns, but may also not belong the original team at all.
2.3 Lost in the continuum

The organizational anonymity problem reveals the difficulties involved in identifying firm boundaries that arises in the more general discussion of the nature of the firm. Arguably, it should make a difference whether the monitor or any other agent are part of one firm or another, but the extent of this difference, if meaningful at all, has yet to be established. As we have seen, Coase’s (1937) dichotomy thesis provided criteria for setting boundaries between firms and markets, but Alchian and Demsetz (1972) challenged this view by arguing that firms are kinds of markets. A related critique of Coase emerged in other key contributions to the contractual theory of the firm during the 1970s and 1980s. In particular, it was increasingly argued on both theoretical and empirical grounds that Coase’s dichotomy “may often be too simplistic,” as Klein, Crawford and Alchian (1978: 326) put it. Barzel (1989: 52) later summed up this position by declaring that Coase’s analysis relies on a “false dichotomy,” that is, it limits the analysis to a binary choice where in fact the set of choices is much greater.  

Richardson (1972: 883), an early proponent of this new view, began by confessing that he, too, “was once in the habit of telling pupils that firms might be envisaged as islands of planned co-ordination in a sea of market relations.” One can only regret such a habit, he continued, since this “simple

---

story, based as it is on a dichotomy between firm and market,” overlooks the existence of a “dense network of co-operation and affiliation by which firms are inter-related.” Firms, Richardson explained, are hardly “islands,” and much of what is usually subsumed under “markets” is actually interfirm cooperation on a more or less long-term basis (Richardson, 1972: 884). It follows that the lens of dichotomy is misleading (Richardson, 1972: 895). Given that it is clearly a mistake to “imagine that reality exhibits a sharp line of distinction” between alternative modes of coordinating economic activity, Richardson (1972: 887) concluded, we would be better served if we realized that “what confronts us is a continuum.”

The “continuum thesis” next appeared in cognate form when Klein, Crawford and Alchian (1978) similarly underlined that short-term contracts of the kind that market transactions are normally associated with are far from the only alternative to employment relations and vertical integration. Indeed, as Klein, Crawford and Alchian (1978: 302) pointed out, various forms of long-term contractual relations are typically among the observed alternatives to the expansion of the firm, “although there is clearly a continuum here.” From this perspective, they reasoned, if unified ownership obtained through vertical integration is to be found at one extreme of the continuum, one can

---

61 In a related discussion, Blois (1972: 253ff) observed that many firms try to develop special relationships with suppliers in order to reap the advantages of vertical integration “without the assuming the risks or rigidity of ownership.”

62 Although the dichotomy thesis was clearly the framework of Markets and Hierarchies, Williamson (1975: 106-109) briefly addressed the issue of intermediate forms of economic organization raised by Richardson (1972) and others. By the end of the decade, Williamson (1979a) proposed a revised framework taking these new aspects into account.
easily picture long-term contracts as being “closer” to the firm along the continuum than short-term contracts.\cite{cohen1979}

Shortly after, Cohen (1979: 587) took the argument an additional step forward by suggesting that there is in actual fact “a range of possible degrees of firm-ness of the contractual arrangement” along the continuum.\cite{cohen1979}

Although in his proposal for research in industrial organization, Coase (1972: 73) had similarly emphasized the need to study “contractual arrangements between firms,” including in this category long-term contracts, leasing, licensing, and franchising, Coase was already aware of the difficulty of drawing boundaries between firms and markets in “The Nature of the Firm.” Of course, Coase (1937: 392, n.1) had observed, “it is not possible to draw a hard and fast line which determines whether there is a firm or not. They may be more or less direction.” Coase therefore accepted, somewhat inconsistently, that there may be potentially numerous situations in which

\cite{cohen1979}

\cite{cohen1979}

Alchian here withdrew one part of the argument made in Alchian and Demsetz (1972), namely that constantly renegotiated short-term contracts characterize both firms and markets. In fact, as Alchian (1984: 38) later explained, “long-term contracts are a necessary attribute of the coalition called a firm.” See also Alchian (2006 [1983]: 264), Alchian and Woodward (1987: 130; 1988: 70), as well as Demsetz’s (1988a: 155-156; 1995: 21, 32) later emphasis on “continuity of association.” Strikingly, this was already Alchian’s (1969b: 348) position: “the firm is a surrogate of the marketplace, but differs in that longer-term general service contracts exist without continuous renegotiations.”

\cite{cohen1979}

Eccles (1981: 342) similarly argued that “pure markets and pure hierarchies are at opposite ends of a continuum of contracting modes.” Likewise, for Milgrom and Roberts (1988: 456), “there is a multidimensional spectrum of institutional arrangements with simple, discrete markets markets and tightly managed hierarchies at two of the extremes.”
the entrepreneur’s authority to direct the actions of others may not fully replace the coordinating role of the price system.\textsuperscript{65} Indeed, Coase (1937: 388) acknowledged that “the degree to which the price mechanism is superseded varies greatly.” It follows that even for the founder of the dichotomy thesis, as Masten (1988: 182) observed, “the distinction between the firm and the market appears to be more a matter of degree than kind, the existence of the firm depending of the amount of direction accorded the manager in the contract.”\textsuperscript{66}

In his article, “The Contractual Nature of the Firm,” Cheung (1983: 2) claimed that the inconsistency in Coase (1937) vanishes once it is realized that his theory of the firm is not exactly a “theory of the firm” but is actually a theory of contractual choice. In fact, Cheung (1983: 3) explained, it is difficult to see Coase in a different light since “Coase d[id] not define ‘the firm’.” Nor did he originally have the clear view of markets, transaction costs and property rights that took shape in the 1960s following Coase’s important contributions (see Coase, 1960). Lacking the appropriate tools, Coase was unable to see the analytical limitations of his own setup. If the later Coase is to be used to save the earlier Coase, Cheung argued, the defective terminology of the latter needs to be discarded. According to Cheung’s

\textsuperscript{65} This seemed to be clear at least in matters of transfer pricing (see Eccles and White, 1988; Bradach and Eccles, 1989; Walker and Poppo, 1991).

\textsuperscript{66} Coase (1990: 11) attempted to overcome this difficulty by replacing his earlier insistence on authority by new considerations regarding the firm’s internal accounting system: “the accounting system … takes the place of the price system of the market.” This idea was not picked up or developed elsewhere since.
(1983: 10, emphasis in original) re-interpretation, “it is not quite correct to say that a ‘firm’ supersedes ‘the market.’ Rather, one type of contract supersedes another type.”

To see why this is the case, it is only necessary to restate how property rights, transaction costs, and contracts are related. Cheung’s straightforward line of reasoning can be readily derived from the basic principles of the economic theory of property rights, namely that “every question of pricing is a question of property rights” (Alchian, 1967: 370), and that “it is the value of the rights that determines the value of what is exchanged” (Demsetz, 1967: 347). On this view, since “the transactions conducted in the market place entail outright or partial transfers of property rights among individual contracting parties” (Cheung, 1969: 23; see also Cheung, 1970: 50), it is clear that the transaction costs incurred in exchange are the costs of establishing and maintaining property rights through “a wide variety of contractual arrangements” (Cheung, 1983: 4). Indeed, across the available “spectrum of alternative contractual arrangements” (Cheung, 1974: 63), the level of transaction costs will depend on the difficulty of pricing or measuring individual behavior.67

Accepting the simple fact that individuals craft contractual arrangements as they see fit, what counts analytically, Cheung (1983: 10; 1987: 56; 1992: 56) continued, is not whether a given contractual arrangement is “a firm,” as Coase had wrongly assumed, but rather how the contractual arrangement deals with pricing or measurement problems. Since every individual is an input owner who enters into various contractual

---

67 A restatement of this view can be found in Allen (1991, 1998).
relations with others, and transfers of rights are more often than not “a matter of degree” (Cheung, 1983: 5), there seems to be no particular reason to reserve the label “firm” to one set of relations to the exclusion of others. Indeed, Cheung (1983: 17) reasoned, “according to one’s view a ‘firm’ may be as small as a contractual relationship between two input owners or, if the chain of contracts is allowed to spread, as big as the whole economy.” In most cases one might wish to consider, Cheung’s (1983: 17) concluded, there are simply too many counter-examples to any proposed definition of “the firm,” and no clear or indisputable answers to questions regarding whether a given contractual arrangement is “one firm or two firms” can be unequivocally provided. Cheung (1992: 56) summed up this position:

Whereas contracts can be separately counted or identified, except for some unusual cases a firm cannot in any economic sense be separately identified. This is because most economic activities in a free-enterprise society are chained by contracts, and as such it is difficult to tell where a firm begins and where it ends. In the usual case, therefore, the size of the firm is indeterminate because there is no cut-off point ... It follows that if we cannot in any meaningful economic sense identify “firms” as separate entities, we do not know what a firm is when we see one in the real world.

Cheung’s argument that it is impossible to provide identity criteria for the firm not only discredited the terminology inherited from Coase’s dichotomy thesis but also overturned Holmström and Tirole’s (1989) concern with Alchian and Demsetz’s story. For Cheung, the irrelevance of the institutional
or legal setting for contracts is a theoretical advantage rather than a source of analytical weakness. From this point of view, Cheung (1983: 17) claimed that tax considerations, business registration procedures and other legal requirements are irrelevant to the economist’s inquiry.

Quite unusually, Cheung explicitly rejected the view that contractual centralization helps identify the firm. Indeed, Cheung never used the expression “nexus of contracts” introduced by Jensen and Meckling (1976: 305-311), regardless of the fact that this “new definition” of the firm does not convey the sharp distinctions that he found so objectionable.68 On the contrary, instead of restricting the analysis to joint production situations, employment relations or any other specific contractual arrangement, Jensen and Meckling (1976: 310) had pointed out that “contractual relations are the essence of the firm, not only with employees but with suppliers, customers, creditors, etc.” It follows, Jensen and Meckling (1976: 311) explained, that “it makes little or no sense to try to distinguish those things that are ‘inside’ the firm (or any other organization) from those things that are ‘outside’ of it.” Employees, from this point of view, are no more “inside” or “outside” the firm than are customers, suppliers, creditors or shareholders.

Like Cheung, Demsetz (1988a: 155) rejected attempts to introduce

---

68 Cheung (1983: 17) pointed out that franchising schemes, subcontracting arrangements, and corporate groups, for instance, all have central contractual agents but are usually viewed as contractual relations between different “firms.” The franchising case was similarly used by Rubin (1978: 232) to deny that there is a “clear-cut division between what is and what is not a firm,” and to argue that “the definition of the franchisee as a separate firm, rather than as a part of the franchisor, is a legal and not an economic distinction.”
various “legal notions of what a firm is and what it is not,” and argued that it is time to realize that “our thinking has been hostage to ‘the firm’” (Demsetz, 1992: 28). Nevertheless, Demsetz (1992: 28, emphasis in original) observed, although “there is no specific set of contractual agreements that defines the firm,” for the purpose of economic analysis, “we can describe sets of contracts as being more or less firm-like once we agree as to what firm-like means.” Contrary to Cheung, Demsetz (1988a: 155ff; 1995: 33ff) turned to the basic Coasean narrative, and suggested that a “firm-like” contractual arrangement must be one that involves a central contractual agent, for in the absence of such an agent the opportunity to economize on transactions would be foregone.⁶⁹ Although both Alchian and Demsetz fully accepted Jensen and Meckling’s label for the central contractual agent (Alchian, 1984: 46; Alchian and Woodward, 1987: 111; 1988: 70; Demsetz, 1988a: 154ff; 1992: 27ff; 1995: 33-34),⁷⁰ Cheung’s main argument remained remarkably unaffected for Alchian.

Indeed, observed Alchian, the striking similarity between families, churches, holding companies, professional sports leagues, cooperatives,

---

⁶⁹ Demsetz (1995: 33-34) compared a situation in which “X enters into a contract with Y, and Y subcontracts parts of the task to Z” with one in which “X contracts bilaterally with both Y and Z,” and concluded that “the second arrangement involv[ing] centrality of contracting [is] of a sort that we associate with firmlike organization.”

⁷⁰ By the late 1980s, the expression “nexus of contracts,” originally associated with the positive agency theory of Jensen, Meckling, Fama and others (see Fama, 1980: 293; Fama and Jensen, 1983a: 302; Meckling and Jensen, 1983: 6; Jensen and Ruback, 1983: 43), was well established in the theory of the firm literature, particularly with its adoption by Alchian and Demsetz.
groups of franchisees, not-for-profit foundations, and so on, is that they are all contractual in essence (Alchian, 1984: 39; 1987: 232; Alchian and Woodward, 1987: 134; 1988: 76). Of course, he continued, there is a sense in which “the firm is distinguished from the other coalitions by its particular objective,” namely the “increased market value of the group’s activities” (Alchian, 1987: 232), but to the extent that this means that a large number of social arrangements fall under the category of “firm,” it is “of little value to try to define the firm as any particular one of them” (Alchian and Woodward, 1987: 134). Ultimately, Alchian (1984: 39) argued, “it makes no difference” (see also Alchian, 2006 [1983]: 268), and the only useful questions are: “why the various types of contractual arrangements are made,” and “what are the consequences” (Alchian and Woodward, 1987: 134; see also Alchian and Woodward, 1988: 76).\(^71\) In this context, Alchian (1987: 234) concluded, “the ‘Firm’ is dead.”

The news of the firm’s death as a useful analytical category spread quickly. In this context, clearly conducive to relaxed definitions, an increasing number of empirical accounts of “quasi-firms” (e.g., Eccles, 1981) and “quasi-integration” (e.g., Monteverde and Teece, 1982) proliferated in the literature.\(^72\) Observing the variety of forms of economic organization

\(^71\) Alchian and Woodward (1987: 134) lamented the fact that “many economists, lawyers, and judges, despite the dictum that 'substance, not form' counts, nevertheless answer a question regarding how an organization ought to be allowed to operate by deciding whether or not the organization ‘is’ a ‘firm’.”

\(^72\) The sentiment that the traditional boundaries of firm were not only difficult to determine but also increasingly porous and “fading” (Picot, Ripperger and Wolff, 1996) grew with the emergence and widespread adoption of the new information and
between the polar cases of the “pure” firm and the “pure” market, many economists accepted Cheung’s (1983: 3) observation that “we do not know what the firm is — nor is it vital to know.” Alchian’s (2006 [1983]: 266) opinion, that “the concept of a ‘firm’ is vacuous and without analytical use,” may very well be correct, they reasoned, since the empirical analysis of specific types of contractual relations in specific industries (e.g., Masten, 1984; Walker and Weber, 1984; Palay, 1984; Anderson and Schmittlein, 1984; Joskow, 1985) could more often than not be carried out without an analytically precise definition of the firm.
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.

Hart (1995: 57-58)

Gibbons (2005) pointed out that a key implication of the continuum thesis is that one necessarily denies that integration between two firms changes anything. In fact, since firms are said to be little more than convenient and ultimately arbitrarily-defined labels for contracting relations, as Gibbons (2005: 231, emphasis in original) explained, “this approach can be seen as denying that integration is anything (besides a label).” Dissatisfaction with the continuum thesis and its various implications pushed the theory of the
firm in various directions, all of which attempted to provide a more satisfactory account of the boundaries of the firm. In an effort to restate the theoretical and empirical importance of firm boundaries, several leading theorists argued that “integration gives greater control (over something),” as Gibbons (2005: 231) put it. The sheer volume of theoretical and empirical work on firm boundaries made it clear that reports of the firm’s death were greatly exaggerated. This chapter presents a synthesis of the three most important attempts to reclaim the boundaries of the firm.

The main approach is Williamson’s (1979a, 1985a) transaction cost economics, that owes much to Klein, Crawford and Alchian’s (1978) hypothesis that vertical integration may be a rational reaction to contractual hazards that arise when parties need to make high levels of “relationship-specific investments.” Significantly, in response to the well-received criticisms addressed by continuum theorists, Williamson proposed to take comparative institutional analysis a crucial step forward by suggesting that asset specificity levels determine the distribution of transactions between markets, interfirm cooperation, and firms (3.1). At the same time, Grossman and Hart’s (1986) dichotomy setup presented a formal model of the firm as a collection of nonhuman assets in order to provide a straightforward view of firm boundaries and the associated allocation of power (3.2). This approach, however, was criticized for its incapacity to explain the nature and boundaries of the new human capital-intensive firm. Addressing this problem, Rajan and Zingales (1998) provided the requisite new foundations for the theory of the firm, based on a more general view of power (3.3).
3.1 The fundamental transformation

As we have seen, in Alchian and Demsetz’s (1972) story the argument that there is no form of power, fiat or authority within firms is based on the assumption of competitive markets in which each party has numerous trading opportunities. Pursuing his favorite comparison, Alchian (1984: 37-38) observed: “If an employee doesn’t care at all whether the firm disappears, because the alternatives elsewhere are just as good – which is the way a customer may regard the presence of a grocery store – then in what useful sense is the employee any more regarded as a member of the coalition than a customer should be so regarded? None at all. What counts is the loss one experiences in the event one must leave the coalition.” For individuals, be they employees or customers, to experience some loss upon exiting a relation, a team or a coalition, there must clearly be something specific about belonging to that relation, team or coalition.

Arguably, the basic setup of competition between interchangeable individuals is profoundly changed when some or all team members choose to specialize, that is, to commit or invest resources that are specific to the coalition. Indeed, each individual’s opportunity set is altered by these “specific investments” that create, according to Klein, Crawford and Alchian (1978: 298), “quasi-rents,” defined as any excess value above their next best use value (see also Alchian, 2006 [1983]: 644; 1984: 37; Alchian and Woodward, 1987: 113). In addition, as Alchian and Woodward (1988: 67) put it, interspecific resources produce a “composite quasi-rent,” namely “that portion of the quasi-rent which depends on continued association with some other specific, currently associated resources.” It follows that for composite
quasi-rents to exist, let alone be maximized, individuals need to accept the sunk or non-recoverable costs of specialization.\footnote{This analysis builds on Becker’s (1975: 26ff) classic distinction between general training and firm-specific training. A concise presentation of the argument is provided in Milgrom and Roberts’s (1992: 329ff) textbook.}

The main problem that arises in this situation is that those contributing to the composite quasi-rent realize that they run the risk of being “held-up” by other team members, as Klein, Crawford and Alchian (1978: 302) explained, building on the terminology introduced by Goldberg (1976a: 439; 1977: 257). Since team members are constantly “competing, even while cooperating, when they act in ways designed to increase their individual shares of the group total” (Alchian, 1984: 36), given the presence of measurement costs, composite quasi-rents are clearly expropriable through “ex post opportunistische behavior.” For instance, some individuals may threaten to withhold effort in order to obtain a greater part of the surplus. Anticipating such problems, rational agents will tend to make “less specific investments [in order] to avoid being ‘locked-in’” (Klein, Crawford and Alchian, 1978: 301), that is, being captive of the hold-up power of others.

Although minimal contractual performance is to some extent ensured by market forces (see Klein and Leffler, 1981), the under-investment problem undermines efficiency, since the opportunity to maximize the joint surplus is foregone. Fortunately, various contractual solutions may help mitigate, albeit imperfectly, this problem. In fact, as Alchian (1984: 37) observed, “anticipation of expropriability of quasi-rent will motivate pre-investment protective contractual arrangements.” From this point of view, individuals
are more likely to become team members by investing in relationship-specific exchange if the protective contractual arrangements designed to deal with the hold-up problem are perceived as “credible commitments,” to use Williamson’s (1983a) expression. Accordingly, Alchian and Woodward (1988: 70) acknowledged that they “can think of neither significant nor interesting cases where teamwork does not create dependencies calling for contractual restraints.”

Vertical integration, the strongest of contractual solutions involving unified ownership, will be used when other contractual restraints, typically embodied in long-term contracts, do not yield the expected results. Klein, Crawford and Alchian (1978: 307) summarized this view with a testable hypothesis: “the lower the appropriable specialized quasi-rents, the most likely that transactors will rely on a contractual relationships other than common ownership,” while “integration by common or joint ownership is more likely the higher the appropriable specialized quasi-rents.” For Alchian (1982a: 238), this is in fact the basic rationale for most if not all of the relevant issues a theory of the firm should address: “‘Expropriable specific quasi-rent’ … is [no] trivial matter [since] the principal rationale for the employer-employee status, for the existence of firms, and perhaps for many

---

74 This is the motivation behind Barzel’s (1987b) argument on capital commitments by the entrepreneur-residual claimant. On Barzel’s (1987b: 103) view, “in order to persuade others to work for him [the entrepreneur] will commit capital of his own to the venture,” and form contracts such that the value of this capital will fall if he reneges on his promise of dividing the surplus according to performance.
other kinds of economic organization, rests on this specificity of investment value.”


Contrary to Alchian’s (2006 [1983]: 262) claim that “without specificity the entity we typically call a firm would not have evolved,” Coase (1988c: 42-43) rejected the view that the main explanatory factors of economic organization are asset specificity and opportunism. Relatedly, Coase also objected to this explanation of what was already becoming the paradigmatic case for studies of contractual hazards and vertical integration, namely the 1926 case of the integration of Fisher Body by General Motors introduced in Klein, Crawford and Alchian (1978: 308ff). Klein (1988: 200ff; 1996: 445ff) subsequently reformulated his explanation, but Coase (2000: 15) again rejected what he called the “prevailing view” on both theoretical and empirical grounds. Interestingly, Klein’s (2000) response made it clear that the debate was really about competing views of the nature of the firm, while Coase (2006) criticized Klein’s empirical work. Although still somewhat active, Coase has not yet reacted to Klein’s (2007) latest response. The General Motors-Fisher Body case has been depicted as one of the “colorful fables” that illustrate key points of economic theory (Casadesus-Masanell and Spulber, 2000: 67), and Spulber (2002: 4ff) has included it in the family of “myths of market failure,” alongside Cheung’s (1973) “fable of the bees” that illustrates externalities, Coase’s (1974b) “fable of the lighthouse” that allegorizes public goods, or Leibowitz and Margolis’s (1990) “fable of the keys” illuminating technological lock-in.
Williamson argued, as specific investments are being made, ex ante large-number bidding markets tend to be transformed into small-numbers bargaining, and eventually into bilateral monopoly situations.\textsuperscript{76} In this context, to the extent that the “organizational imperative” (Williamson, 1985a: 32) is to safeguard transactions from the contractual hazards that arise with the possibility of opportunistic behavior,\textsuperscript{77} special “governance structures” – defined by Williamson (1979a: 239) as “institutional matrices” in which transactions are negotiated and executed (see also Williamson, 1996: 378) – will tend to supplant standard market exchange when “transaction-specific values are great” (Williamson, 1979a: 244-245).\textsuperscript{78}

Arguably, Williamson hypothesized, efficiency is enhanced if “transactions, which differ in their attributes, are aligned with governance

\textsuperscript{76} See Williamson (1975: 27) for an earlier discussion of small number situations, and Williamson, Wachter and Harris (1975: 256ff) for a preliminary discussion of “task idiosyncracies.”

\textsuperscript{77} On Williamson’s (1979a: 234) view, “opportunism is a central concept in the study of transaction costs,” and indeed “the world of contract in the absence of opportunism is uninteresting” (Williamson, 1998a: 704). Williamson (1985a: 51, n.8; 1993c: 101) justified his use of the broader term “opportunism” as opposed to the more common term “moral hazard” by the fact that opportunism is a less technical term that includes both ex ante adverse selection and ex post moral hazard, and is relevant at the recontracting stage: opportunistic behavior is the more general category when considering “incomplete contracting in its entirety” (Williamson, 1990d: 189; 1996a: 9; 2005: 388).

\textsuperscript{78} It is clear that parties will increase specificity levels only if the governance costs are more than offset by the increased value of additional specificity levels. This is a variation on the Demsetz thesis.
structures, which differ in their costs and competencies, in a discriminating (mainly, transaction-cost-economizing) way” (Williamson, 1991: 277; see also Williamson, 1998b: 77; 1993a: 40; 2000: 599; 2002b: 441; 2005a: 378; 2005b: 5). With this “discriminating alignment hypothesis,” that involves “dimensionalizing” and matching transactions and governance structures, Williamson took comparative institutional analysis a crucial step forward in line with Simon’s (1978: 6) observation that, “as economics expands beyond its central core of price theory … its central concern [becomes] a much more qualitative institutional analysis, in which discrete structural alternatives are compared.” In the process, Williamson also responded to the concern that the firm-market dichotomy may be a false dichotomy. Consequently, his own earlier dichotomy analysis was revised to include, “between” markets and hierarchies, a large variety of contractual arrangements and “intermediate modes of organization” (Williamson, 1979a: 234), later dubbed “hybrid forms” (Williamson, 1988a: 88; 1991: 269).79

Williamson introduced an innovative argument to support his claim that “each viable form of governance – market, hybrid, and hierarchy – is defined

79 Contrary to some continuum theorists who rejected the usefulness of the concept of “firm” as an analytical category, Williamson argued that the distinction between firms and markets remains central. Indeed, the very notion of “hybrids” implies the analytical usefulness of the polar forms (see Williamson, 1975: 109). Contrary to his earlier contention that “transactions of the middle kind” are very difficult to organize and hence tend to be unstable, Williamson gradually became persuaded that such transactions are quite common (see Williamson, 1985a: 83). See Ménard (2009) on the evolution of Williamson’s theory of hybrid forms, and Hodgson (2002) for a critical discussion of the notion of “firm-market hybrid.”
by a syndrome of attributes that bear a supporting relation to one another” (Williamson, 1991: 271). He pointed out that some of the confusion in the literature is imputable to the fact that the notion of contract is employed too loosely by economists, and that a “deeper understanding of the nature of contract” (Williamson, 1979a: 235) was necessary. Turning to the legal literature, Williamson adopted Macneil’s (1978) three-way typology of contract law, in which each type of contract law is characterized by distinct properties of stability, flexibility, and remedy possibilities, and suggested that this typology can be used to construct a framework in which firms are neither simply kinds of markets, like in Alchian and Demsetz’s (1972) account, nor indeterminate and unimportant, as continuum theorists such as Cheung (1983) had argued. As Williamson (1991: 270) put it:

Firms are not merely extensions of markets but employ different means [and] discrete contract law differences provide crucial support for and serve to define each generic form of governance.

Williamson began his demonstration by observing that three dimensions of transactions tend to increase transaction costs (Williamson, 1979a: 239; 1981a: 676; 1981b: 555; 1985a: 52ff; 1985b: 181; 1986b: 179ff; 1988a: 70). Alongside the self-explanatory first dimension, transaction frequency, an important second dimension is uncertainty, of both the external “state-
“contingent” and internal “behavioral” or “strategic” kinds (see Williamson, 1985a: 56-60). Regardless of the source of uncertainty, given that many key contracts tend to be incomplete because the contracting parties are unable to conclude “comprehensive” or “complete state-contingent contracts,” a key problem that designers of governance structures need to deal with is adaptation to typically unforeseeable changes.81 The most critical dimension, however, is the third. As Williamson (1988a: 70) clarified, following Klein, Crawford and Alchian (1978), “most of the refutable implications of transaction cost economics” hinge on the degree of asset specificity (see also Riordan and Williamson, 1985: 366).82

Having thus dimensionalized transactions, Williamson (1979a: 248) went on to argue that the “main governance structure for non-specific transactions of both occasional and recurrent transacting” is “market governance” (see also Williamson, 1985a: 73). Spot markets are the loci of “truly discrete exchange” in that each transaction, as Macneil (1978: 856) had explained, is “entirely separate not only from all other present relations but from all past and future relations as well.” Crucially, the identity of parties is irrelevant, contingencies are fully specified, contracts are

81 Williamson (1991: 277-278) credited Barnard (1938: 6) and Hayek (1945: 523) for underlining “adaptation as the central economic problem.”

82 Williamson (1991: 281) identified six forms of asset specificity raising the costs of investment redeployment: site specificity, physical asset specificity, human asset specificity, brand name capital, dedicated assets, and temporal specificity. In earlier work Williamson (1983: 526; 1985a: 55) had identified four different types of asset specificity: site specificity, physical asset specificity, human asset specificity, and dedicated assets. According to Williamson’s (1996: 377) glossary, these are the most common.
interpreted and executed in a strictly legalistic way, litigation is used to solve disputes, and market alternatives offer protection against opportunism. Discrete transactions of this kind, governed by what Macneil (1978: 856ff) called “classical contract law” (Williamson, 1979a: 236; 1985a: 69; 1991: 271), are the paradigmatic focus of the standard economic model (see Goldberg, 1976b: 49, Macneil, 1978: 862ff; 1981: 1020). However, as Williamson (1979a: 235) pointed out, “many contractual relations are not of this well-defined kind.” It follows that in many cases the “fiction of discreteness” (Williamson, 1979a: 238) needs to be replaced by more realistic analytical categories

In accordance with Macneil’s (1978: 865) own view that “the discrete transaction is at one end of a spectrum [while] at the other end … we … come to the firm itself,” Williamson argued that departures from market governance will be observed for recurring transactions underpinned by increasing levels of specific investments (Williamson, 1979a: 250; 1983a: 537; 1985a: 75; 1985b: 185; 1986a: 85ff; 1991: 272ff; 1996: 16). As specificity and mutual dependence levels build up, he explained, market alternatives provide only limited protection from opportunism because identity plainly matters, as does the continuity of the relation. But for “bilateral structures” typical of long-term interfirm cooperation and similar hybrid forms to replace market governance (Williamson, 1979a: 250; 1991: 271), “special adaptive mechanisms” are needed “to … restore efficiency when beset by unanticipated disturbances” (Williamson, 1991: 272) that may lead to a dispute. Commitment to third-party assistance, or arbitration, is one such special adaptive mechanism that safeguards specific investments (Williamson, 1983: 527-528; 1991: 272; 1996: 131-132).
Contractual safeguards of this kind are typical of Macneil’s (1978: 865ff) description of “neoclassical contract law.” A key aspect of the neoclassical contracting regime, according to Williamson (1991: 272), is that Llewellyn’s (1931: 737) notion of “contract as framework” applies. Contrary to the strictly legalistic classical contract law, in other words, neoclassical contracts are much more elastic in the sense that they are compatible with flexible interpretations of contractual clauses. Since the spirit of the contract counts as much as the letter of the contract, neoclassical contracts allow a significantly greater degree of adaptation. Despite such adaptability, however, neoclassical contracts are not indefinitely elastic. As unforeseen disturbances become highly consequential, the costs of maintaining the long-term contractual relation via arbitration rise due to the fact that the bilateral governance of interfirm agreements still involves interactions across a market interface, and that “the autonomous ownership status of the parties continuously poses an incentive to defect” (Williamson, 1991: 273).

This situation calls for an even more elastic and adaptive arrangement, one that would eliminate the incentive to defect altogether. For Williamson (1979a: 252ff; 1985a: 78ff), “unified governance,” namely unified ownership and internal organization through vertical integration, qualifies as the requisite elastic and adaptive mode of organization. A crucial advantage of the firm as a “unified governance structure” (Williamson, 1985a: 73), he

---

83 Williamson’s use of Macneil’s typology changed over time, as did his typology of governance structures. For the present purposes, some aspects of Williamson’s (1979a) initial discussion, for instance the concepts of “trilateral governance” and “relational contracting,” are disregarded because both these notions have disappeared from Williamson’s account.
observed, is that “adaptations can be made in a sequential way without the need to consult, complete, or revise interfirm agreements” (Williamson, 1979a: 253). Indeed, Williamson (1991: 274) argued, “internal organization, hierarchy, qualifies as [the most] elastic and adaptive mode of organization,” because “the implicit contract law of internal organization is that of forbearance,” meaning that courts typically refuse to hear technical disputes between internal divisions of the firm, and that internal procurement contracts are not legally enforceable. Instead of litigation (classical contract law of market governance) and arbitration (neoclassical contract law of hybrid governance), then, “hierarchy is its own court of ultimate appeal” (Williamson, 1991: 274), and disputes are solved by internal “administrative controls” (Williamson, 1991: 280).

In this manner, the organizational anonymity problem appears to have been resolved because “whether a transaction is organized as make or buy … matters greatly in dispute-resolution respects” (Williamson, 1991: 275). Indeed, the institutional setting for contracting matters, and there is a clear sense in which intra-firm fiat is underpinned by the institutional setting. This, according to Williamson (1991: 276), is precisely what Alchian and Demsetz

---

84 Forbearance doctrine, as Williamson (1991: 274-275) explained, is akin to but more general than the “business judgment rule” according to which corporate directors act on an informed basis in the firm’s best interest. Supporting Williamson’s view is Manne’s (1967: 271) argument that the rule creates a “quasi-jurisdictional barrier to prevent courts from exercising regulatory powers over the activities of corporate managers,” and Gilson’s (1986: 741) opinion that “the courts’ abdication of regulatory authority through the business judgment rule may well be the most significant common law contribution to corporate governance.”
(1972) ignored when they argued that there is no meaningful difference between the relation between a grocer and his employee, and the relation between a grocer and his customer. For Williamson, prior neglect of different types of contract law doctrines, but also prior lack of consistent dimensionalization of transactions, explain Alchian and Demsetz’s error. Interestingly, Williamson (1991) put forward this new account of the sources of intra-firm authority in lieu of his earlier view of intra-firm authority based on the employment relation (e.g., Williamson, 1975; Williamson, Wachter and Harris, 1975), perhaps because he came to agree with Coase (1988c) that identifying employment relations and the firm is unnecessarily restrictive.

Contrary to the view that firms are extensions of markets, which seems to imply that the incentive structure within firms should be the same as that to be found in markets, Williamson (1991: 275) further identified the difference between firms and markets as a difference in incentive intensity: “as compared with markets, internal incentives … are flat or low-powered, which is to say that changes in effort expended have little or no immediate effect on compensation.” But instead of viewing this as the problem to be resolved, as in Alchian and Demsetz’s setup, from Williamson’s (1985a: 140ff; 1991: 275) point of view low-powered incentives are an advantage of the firm over the market because they help elicit greater intra-firm cooperation. That employees lack the high-powered incentives typically associated with residual claimant status (see Williamson, 1985a: 132) is beneficial precisely because the incentives to defect are muted.

The generalization of the preceding analysis to all sorts of constituencies, including creditors or customers (Williamson, 1984: 1206ff; 1988b: 574-575; 2008: 250), where “the very same contractual apparatus is
uniformly applied to each constituency” (Williamson, 1985a: 318), implies the presence of a central contractual agent. Arguably, Williamson (1985a: 318) observed, it is “the centrality of management” that “distinguishes it from all other constituencies.” As Williamson (1985b: 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” (see also Williamson, 1984: 1225). Significantly, Williamson accepted an important aspect of Alchian and Demsetz’s theory, namely that management’s superior knowledge is the basis of superior decisions about efficient or profitable resource combinations.

Clearly, Williamson also accepted the basic idea of the “nexus of contracts” developed by Coase’s followers although he expressed some reservations about the expression itself, suggesting at one point its replacement by the less legalistic expression, “nexus of treaties” (Williamson, 1990a: 3ff).\textsuperscript{85} This suggestion did not attract much enthusiasm, and even Williamson barely mentioned it in subsequent work. Nevertheless, from Williamson’s account it is clear that the notion of nexus of contracts does not automatically imply Jensen and Meckling’s (1976) rejection of firm boundaries. Contractual relations are within firm boundaries to the extent that they depend on the ability of central management to resolve disputes internally without appealing, and without being able to appeal, to an external

\textsuperscript{85} It is possible that Williamson had in mind something similar to Macneil’s (1978: 901) view of the firm as a “minisociety.” See Adelstein (2010) for this interpretation. In a recent interview, Williamson (2007: 382) emphasized the expression’s relation to Llewellyn’s notion of contract as framework.
dispute-resolution mechanism. Arguably, incentive intensity, low within firms and high outside, is a matter of degree, while forbearance uniquely defines the nature of the firm in a way that even administrative controls cannot for their presence, too, is a matter of degree. This means that employees, including managers, are inside the firm, contrary to suppliers and customers. Furthermore, creditors are also on the outside while equity providers are within the firm (Williamson, 1984: 1228; 1988b: 580).

### 3.2 The meaning of ownership

Williamson has often distinguished his theory from alternative approaches to economic organization. In particular, Williamson (1985a: 24; 1985b: 188; 1989: 149; 2000: 610) has repeatedly contrasted the “measurement branch”

---

86 A formal model of this idea can be found in Wang and Zhu (2004).

87 Williamson (1988b: 576) likened debt to outside procurement and equity to vertical integration. As Williamson (2008: 248) recently explained, debt and equity are not simply alternative financial instruments but also alternative modes of governance.

88 Williamson (1990b: 62) clearly separated the incomplete contracting world of transaction costs economics from the complete contracting tradition (e.g., agency theory). Within the new institutional economics, Williamson (2000: 596-600; see also Williamson, 1993a) distinguished transaction cost economics’s focus on governance, that is, on the “play of the game,” from other theories that focus either on the higher analytical level of the institutional environment, that is, on the “rules of the game” (e.g., positive political economy), or on the analytically lower level of incentive alignment (e.g., agency theory).
of transaction cost economics developed along the lines of Alchian and Demsetz (1972), and the “governance branch” to which he is the foremost contributor. Alchian and Woodward (1988: 69-70) reformulated this distinction by metaphorically observing that “early explorers of transaction costs” can be divided into the “measurement expedition” and the “asset specificity expedition.” As we have seen, the latter expedition set out to reclaim firm boundaries by focusing on how contractual relations are fundamentally transformed once parties make, or need to make, relation-specific investments. The main prediction is that transactions tend to shift out of markets into hierarchies through vertical integration in order to mitigate hold-up risk when expropriable quasi-rents are high (Klein, Crawford and Alchian, 1978; Williamson, 1979a, 1985a).

A pioneering formal model of this setup was introduced by Grossman and Hart (1986), and further developed in Hart and Moore’s (1990) “Property Rights and the Nature of the Firm.” Although Hart (1995: 87) later claimed that “the hold-up problem is a useful [but] not an essential part of the approach,” the so-called “new property rights theory” of the firm is an

89 The approach was initially welcomed by Williamson (1985a: 136, n.4; 1990b: 68, n.15), who depicted it as belonging to the “formal” stage of the “evolving science of organization,” that the preceding “semiformal” stage of his own transaction cost theory had made possible (Williamson, 1993a: 41-42, 55; 2010a: 686). However, although Williamson (2010a: 686) recently restated this position, he has emphasized the numerous substantive differences between the two setups (Williamson, 2000: 605ff; 2002a: 188). Most commentators today underline these differences (see Holmström and Roberts, 1998: 75; Foss, 2000: xlii; Whinston, 2003: 4; Garrouste, 2004: 371; Gibbons, 2005: 201).

In addition, Hart (1988: 121) explained, like Coase, Williamson vaguely reduced the costs of integration to increased bureaucracy, and failed to recognize that the costs and benefits of integration are “two sides of the same coin” (Hart and Moore, 1990: 1120). This lacuna complicates the task of countering the claims made by continuum theorists who deny that integration produces real effects and “sidestep the issue entirely,” as Moore (1992: 494, n.2) put it, “by arguing that everything is contractual, and that firms are a mirage.” For Hart (1989: 1764-1875), this position is plainly mistaken “given that mergers and breakups occur all the time, and at considerable transaction cost.” Since it seems unlikely that such changes are “cosmetic,” he continued, the theoretical challenge is to explain how integration produces “real effects on incentives and opportunistic behavior” (see also Hart, 1990: 696).

---

90 Regardless of the ongoing debate between Coase and Klein about the events that led to the integration of Fisher Body by General Motors, Hart has often used the example to illustrate the property rights theory of the firm (see Hart, 1988: 125; 1989: 1767-1773; 1995a: 7, 23-33; 1996a: 372-375).
On Grossman and Hart’s (1986: 693) view, the problem with existing theories is that they lack “a sufficiently clear definition of integration.” This is what makes it hard to decide “whether [these theories] are designed to explain the types of people called employees or instead the types of assets under the control of a single ownership unit” (Grossman and Hart, 1986: 693). Existing theories, in other words, are insufficiently clear on what exactly is acquired through integration. Addressing this problem, Hart (1995: 29) observed:

Imagine that firm A acquires firm B … What exactly does firm A get for its money? At least in a legal sense, the answer seems straightforward: firm A acquires, i.e. becomes the owner of, firm B’s assets … Excluded are the human assets of those people working for firm B; given the absence of slavery, the human capital of these workers belongs to them both before and after the acquisition.

The distinction between alienable nonhuman assets and inalienable human assets, defined as the “skills of … workers and managers” (Hart, 1996a: 373), is central in the property rights theory of the firm. Arguably, only nonhuman assets, including “hard” assets such as machines and buildings, as well as “soft” assets such as contracts and patents, literally change hands in a merger. It follows, Hart (1995: 56) conjectured, that “(at least some) nonhuman assets are an essential feature of a theory of the firm.”

91 In fact, if

91 Nonhuman assets include machines, inventories, buildings or locations, cash, client lists, patents, copyrights, contractual rights and obligations, and so on (see Hart, 1989: 373).
the firm is to be meaningfully viewed as a single ownership unit, a case can be made for the fact that the firm itself is identical with these assets. From this point of view, Hart (1989: 1771, n.48) unequivocally stated that “in the language of the property rights approach, ‘firm’ is shorthand for a collection of assets” (see also Grossman and Hart, 1986: 692-293; Hart, 1990: 696; Hart and Moore, 1990: 1119-1120; Moore, 1992: 494).92

This definition allows one to make sense of the fact that firms, like other assets, are bought and sold in markets – a fact that seems at odds with the view that firms are teams or coalitions of human beings. Furthermore, this definition makes Cheung’s perspective easy to refute, since “firms appear easy to identify” (Moore, 1992: 494, emphasis in original). Indeed, it is easy to ascertain firm boundaries: nonhuman assets that belong to the firm are “inside” firm boundaries while nonhuman assets that do not are “outside.”93 While these advantages are somewhat tautological, proponents of the collection of assets view argued that their setup contains important insights for the contractual theory of the firm. In Hart’s words, the property rights setup “opens the door to a theory of ownership” (Hart, 1988: 123) that

92 This definition is closely related to the corporate finance literature. Myers (1977: 171), for instance, argued that “at any point in time the firm is a collection of tangible and intangible assets.”

93 According to Hart (1995: 62), this view holds equally for small firms and for large corporations. Commenting on this approach, Langlois (1998: 16, n.18) pointed out that few in the literature have been able to define the boundaries of the firm as “criply and consistently.” See also Masten (1998: 58) and Foss (2000: xlii).
can “throw light on the meaning of authority” (Hart, 1996a: 371) such that “the Alchian-Demsetz observation that an employer has no more power over an employee than one independent contractor has over another … is no longer true” (Hart, 1996a: 377).

Hart argued that both Coase’s (1937) theory of authority and Alchian and Demsetz’s (1972) objection, that employees cannot be forced to do what an employer wants, are flawed for the same reason. Since the distinction between nonhuman assets and human assets is missing, the debate is misleadingly focused on authority over human assets, instead of acknowledging that, just like ownership applies to nonhuman assets only, so too is “authority … best interpreted as applying to non-human rather than human assets” (Hart, 1996a: 383). In fact, as Hart and Moore (1990: 1150) observed, “in the absence of any nonhuman assets, it is unclear what authority or control means.” By contrast, when nonhuman assets are present, it is not difficult to understand that “an employer indirectly gains ‘authority’ over an employee as a result of owning an important asset” (Moore, 1992: 498), and that, generally speaking, “authority over assets translates into authority over people” (Hart and Moore, 1990: 1150).

To see why this is the case, Hart (1996a: 375) explained, it is important to recognize that the notion of authority is meaningful only insofar as contracts contain gaps that need to be filled in “as the future unfolds” (Hart, 1995: 24). Indeed, there is no need to have a “mechanism by which the gaps are filled in as time passes” (Hart, 1988: 123) if contracting costs are zero,

---

94 Tirole (1988a: 464) made a similar point.
and parties are able to write complete state-contingent contracts. In the real world, however, when parties have to deal with unforeseen contingencies in which nonhuman asset uses have not been specified, perhaps because they are non-contractible, authority is effectively exercised by the party that holds the “residual rights of control” over asset uses (Hart, 1996a: 376). Since “possession of residual control rights is taken virtually to be the definition of ownership,” Hart (1995: 30) continued, authority resides with asset owners who hold all the economically significant ex post decision rights that have not been specified in the initial contract, and that do not violate some law or custom (see also Hart, 1989: 1765; 1900: 696; 1996b: 28-29; 1998: 331).

Hart and Moore (1990: 1150) argued that a key residual control right is the “ability to exclude people from the use of assets.” Alchian and Demsetz had acknowledged that the owner can unilaterally change team membership, but had failed to grasp that exclusion rights give employers leverage or indirect authority over employees. For Hart (1996a: 377), “an employee is more likely to do what the employer wants than a grocer is to do what the customer wants, because the employer can deprive the employee of the assets the employee needs to be productive, whereas the customer cannot deprive

---

95 This observation is the starting point for a large variety of incomplete contracting models that extend well beyond the theory of the firm to issues ranging from capital structure and industrial organization to privatization and international trade. Complementary overviews are provided by Nicita and Pagano (2005), and Aghion and Holden (2011).

96 Hart (1996a: 383) claimed that the Weberian distinction between “formal” (de jure) and “real” (de facto) authority discussed in Aghion and Tirole (1997) does not change the thrust of the argument.
the grocer of the assets the grocer needs to be productive” (see also Hart, 1989: 1771; 1995a: 58; Moore, 1992: 496). At the same time, as Hart (1996a: 377) pointed out, “whereas a grocer can fairly easily find another customer to serve, an employee typically cannot.” Since an employee’s outside options are typically slim given the lack of asset ownership, he reasoned, “the employee is more likely to pay attention to the employer than a grocer is to the customer because the employee’s livelihood is at stake.”

This argument holds even in the absence specific investments, as does the fact that the allocation of residual control rights helps define the firm as a single ownership unit: employers hold control rights over non-specified actions of their employees because employers own the assets involved, whereas contractors do not hold control rights over non-specified actions of contractees or their employees because contractors do not own their contractees’ assets. On the other hand, Alchian and Demsetz’s definition of ownership as rights to residual income is comparatively less successful. Employees who receive residual income rights through various incentive schemes are not “owners” with “authority,” according to Hart (1988: 125; 1996a: 30), and profit-sharing schemes between independent firms are quite

97 Alluding to the debate with the radicals, Hart (1995: 5, n.5) recognized that the new property rights approach shares with Marxian theories of the capitalist-worker relationship “the idea that an employer has power over a worker because the employer owns the physical capital the worker uses (and therefore can appropriate the worker’s surplus).” On this view, Hart (1996a: 379) conceded, “an employer’s authority is represented not by the ability to force an employee to do what s/he wants, but rather by the ability to obtain a substantial share of the ex post surplus from the relationship through the control of non-human assets.”
possible. Clearly, the divisibility of residual income rights contrasts with the indivisibility of residual control rights: it is “difficult if not impossible to allocate 80 per cent of the residual control rights to party A and 20 per cent to party B” (Hart, 1995: 64, n.13).\(^9\)

These considerations are important when parties in a buyer-supplier relationship are required to make complementary specific investments (e.g., buy specialized equipment, or learn to use each other’s assets) in order to maximize the joint surplus over time. Both parties realize that once there is “something binding the partners together” (Hart, 1995: 26), the division of the joint surplus in the event of unforeseen contingencies will depend on each party’s bargaining power rather on the initial contract, “best seen as providing a suitable backdrop or starting point for renegotiation” (Hart, 1996a: 373).\(^9\) Since bargaining power depends on “the position of each party if the other party does not perform (e.g. if the other party behaves

\(^9\) Nevertheless, for Hart (1995: 64-66; 1996b: 30-32; 1998: 334), complementarities between control rights and income rights make it sensible to bundle these rights. The result is therefore similar to Alchian and Demsetz’s solution to the monitor shirking problem.

\(^9\) Departing from their original renegotiation model (Hart and Moore, 1988; see also Hart and Tirole, 1988) in recent work, Hart and Moore (2007, 2008) and Hart (2008a, 2009a) focus on how the pre-fundamental transformation contract becomes a “reference point” for subsequent haggling based on new behavioral considerations: parties withhold cooperative effort when “aggrievement costs” (Hart and Moore, 2007: 184; 2008: 4ff; Hart, 2008: 406ff) are positive, that is, when parties feel that the spirit of contract has not been respected. Like in Klein’s (1996, 1998) account, this form of hold-up behavior does not occur within a certain “self-enforcing range” of states of the world. See Fehr, Hart and Zehnder (2009) for experimental evidence.
opportunistically),” Hart (1995: 3-4) explained, it may be optimal for the party most exposed to hold-up risk to “purchase all the rights except those specifically mentioned in the contract” (Grossman and Hart, 1986: 692). In fact, Hart (1998: 331) hypothesized, asset ownership “will tend to gravitate to the person for whom residual control rights are particularly important.”

Vertical integration, appropriately defined as “the purchase of the assets of the supplier (or a purchaser) for the purpose of acquiring the residual rights of control” (Grossman and Hart, 1986: 716; 1987: 535), will clearly have real effects by changing the “parties’ outside options and their incentives to engage in hold-up” (Hart, 2009: 280). By acquiring firm B, the incentives of firm A’s owner to engage in relationship-specific investments improve since integration increases firm A owner’s bargaining power and share of ex post surplus by eliminating the hold-up risk that exists as long as firm B remains independent (see Hart, 1995: 32). Symmetrically,

\[\text{\textsuperscript{100}}\text{For Hart (1996a: 373), “the more assets a party owns or controls, the stronger … its position if the relationship breaks down and … the stronger … its bargaining position in any ex post renegotiation.” In Grossman and Hart’s (1986: 704) terminology, ownership increases bargaining power by raising the owner’s “status quo utility” in the event of no renegotiation. Holmström and Tirole (1989: 69) pointed out that on this view “ownership defines the default options in an incomplete contract.”}\]

\[\text{\textsuperscript{101}}\text{From this perspective, Hart’s (1995a: 44ff) discussion of optimal asset ownership structures is based on the idea that assets will tend to be acquired by those to whom they are indispensable, or when they are highly complementary to the assets already owned (see also Hart and Moore, 1990: 1123-1124).}\]

\[\text{\textsuperscript{102}}\text{It follows that the value of residual control rights determines the value of the acquired firm.}\]
given that firm B’s previous owner is now firm B’s manager and employee of
firm A’s owner, his incentives are comparatively weaker. Arguably, as
Moore put it, this is the trade-off to be evaluated “primarily in terms of the
(aggregate) effects on the incentives” (Moore, 1992: 497) that can explain
“the forces that determine whether transactions are conducted ‘within the
firm’ as opposed to ‘through the market’” (Moore, 1992: 494, emphasis in
original).

Hart (1989: 1772; 1995a: 56ff) summarized the main insights of the
property rights theory of the firm in his response to Klein’s (1988: 208)
variation on the Hobbesian question: what, Klein had asked, prevents firm
B’s employees from manifesting their concern about the merger by quitting
en masse, and announcing one morning that they have become a new firm? A
collective hold-up threat of this kind can be credible, Hart (1995: 57)
observed, only in the absence of some sort of “glue” holding firm B
employees in place, that is, only when there there is no ultimate source of
firm B value “over and above the workers’ human capital.” Klein’s concern
is therefore justified regarding his own theory of vertical integration as the
purchase of “organizational capital … embedded in the human capital of the
employees” (Klein, 1988: 208), and Hart (1988: 137) pointed out the same
weakness in Kreps’s (1990a: 766ff) notion that the firm is composed entirely
of “reputational capital” held together by “reputational glue” (Kreps, 1990a:
762; 1990b: 93).

The property rights view fares much better in this respect, Hart (1995:
57) observed, since it is easy to demonstrate that “a firm’s nonhuman assets
... simply represent the glue that keeps the firm together.” Indeed, when the employees’ human capital is highly complementary to the nonhuman assets that change hands in the merger, and outside options are slim, firm B’s employees accept firm A’s authority because firm A now holds the residual control rights to key nonhuman assets, such as a physical location, the firm’s name, its distributional network, client lists, outstanding contracts (including employees’ non-compete clauses), and so on. In Hart’s (1995a: 58-59) words:

If such assets do not exist, then it is not clear what keeps the firm together, or what defines authority within the firm. 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 discussion reveals an important consequence of the definition of the firm as a collection of nonhuman assets, namely that there is an important sense in which employees are not really part of the firm. Indeed, the firm is fundamentally composed of an “owner” (perhaps representing a coalition of owners) and the collection of nonhuman assets under the owner’s control. The owner, then, is the central contractual agent with whom employees, suppliers, creditors, and all other constituencies, contract as outside parties. In this sense, the new property rights theory of the firm is similar in spirit to Alchian’s (1984: 39) conception of the contractual nexus, according to which

103 Even if reputation is viewed as a tradeable asset (see Tadelis, 1999), it is still very difficult, according to Hart (2001: 1714), to account for the “stickiness in the firm or system, so that a firm’s reputation can be separated from that of key personnel.”
“a firm is a coalition of interspecific resources owned in common, and some generalized inputs,” where the most notable generalized inputs are employees whose services are “rented” by the firm owners (see also Alchian and Woodward, 1988: 70; Alchian, 2006 [1978]: 644). This is also Fama’s opinion, although for Fama (1980: 290) “ownership of the firm is an irrelevant concept.”

### 3.3 In search of new foundations

In their review of the literature, “The Boundaries of the Firm Revisited,” Holmström and Roberts (1998: 79) claimed that the property rights view is hardly as successful as its proponents make it out to be. On the contrary, they argued, the emphasis on bargaining and exit rights is better suited to explain entrepreneurial incentives in markets rather than firms. Since the property rights view is really about the distribution of nonhuman asset ownership among “representative entrepreneurs,” as Holmström (1999: 100) similarly explained, the approach is silent on internal organizational issues, and “seems of little empirical relevance” (Holmström and Roberts, 1998: 79). Importantly, Holmström (1999: 87) continued, “the same critique that was directed at Alchian and Demsetz’s vision of the firm, that organizational affiliations did not matter for transactions, could be directed at the Hart-Moore model just as well. Individual ownership of assets does not offer a theory of organizational identities unless one associates individuals with
firms.”104 The property rights view, Holmström (1999: 75) concluded, “despite its express objective to explain the boundaries of the firm, fails to do so.”

Holmström (1999: 75) argued that the way forward lies in a synthesis of the property rights definition of firm boundaries based on asset ownership and a measurement costs explanation of internal organization built on Alchian and Demsetz’s (1972) insights.105 In addition, according to Holmström and Roberts (1998: 92), for the theory to “prove useful for empirical studies,” it is necessary to develop the idea that asset ownership is not the only source of incentives, and to explain how various substitutes for ownership elicit employee cooperation within the low-powered incentive structure of the firm. The firm, explained Holmström and Milgrom (1994: 990), should be properly viewed as a “coherent incentive system” (see also Holmström, 1999: 89) within which incentive pay, job design, task assignment, and other complementary instruments such as transfer pricing schemes help internalize the “contractual externalities” that arise in the presence of measurement costs (Holmström and Tirole, 1991: 215; Holmström and Roberts, 1998: 86ff; Holmström, 1999: 76, 95ff).106

104 Kreps (1990b: 99) likewise observed that the property rights view “is not a theory of the firm per se” since “capital ownership by single entrepreneurs is as likely a consequence as is ownership by an entity with a firm’s legal status.”

105 Hart and Moore’s (2005) model of the design of hierarchies, viewed as the division of labor between specialization and coordination, did not include measurement cost considerations.

106 These features are nevertheless absent in Hart and Holmström’s (2010) latest
A related critique of the property rights approach is behind Rajan and Zingales’s (1998, 2000, 2001a, 2001b) addition to the asset specificity expedition. The empirical irrelevance of Hart’s theory, they argued, is not as severe as Holmström and Roberts had claimed. Of course, taken literally, “the property rights view of the firm … applies only to the entrepreneurial firm,” acknowledged Zingales (1998: 501), but it is easy to see that Hart’s explanation of the employer’s power over employees can be straightforwardly extended to Chandler’s (1977: 1ff) “modern business enterprise” typical of managerial capitalism.\(^{107}\) Vertically integrated, nonhuman asset-intensive hierarchies of this kind, Rajan and Zingales (2000: 209) explained, were “well defined by the ownership of assets,” because nonhuman assets were the “main source of value and control” over employees (Zingales, 2000: 1643). Given the absence of competition in the intermediate products market and the unavailability of finance, the lack of outside options literally glued employees to the firm’s nonhuman assets.

However, as Rajan and Zingales (2001a: 207) emphasized, this form of glue is now “evaporating.” Indeed, “vertically integrated organizations that enjoy rents because their [nonhuman] assets … give them an unassailable position in the industry are becoming creatures of the past” (Rajan and discussion of firm boundaries based on the “contracts as reference points” framework.

\(^{107}\) The “modern industrial enterprise,” as Chandler (1990a: 3ff) also referred to it, was the object of the managerial theory of the firm developed by Baumol (1959), Marris (1964), and Williamson (1964), and appears in the bulk of the industrial organization, corporate finance, and corporate governance literatures. Williamson (1974, 1981c, 1985a) provided a transaction cost economics rationale for the “modern corporation” (see also Williamson and Berkovitz, 1996). See Mueller (2003) for a recent overview.
Zingales, 2000: 212). In fact, the “breakdown of the traditional vertically integrated firm” (Rajan and Zingales, 2000: 213) has coincided with “major technological, regulatory, and institutional changes [that] have made finance more widely available” (Rajan and Zingales, 2001a: 206). Barriers to and costs of entry have considerably dropped at the same time as global trade has expanded the size of the market, producing a highly competitive and noisy environment in which firms need to continually invest in brand name distinctiveness and innovate in order to survive. In this context, Rajan and Zingales observed, since “the innovative energy of employees has to be harnessed in making the firm more creative and productive” (Rajan and Zingales, 2004: 94), inalienable “human capital has replaced [alienable] capital as the main source of value” (Rajan and Zingales, 2000: 224).

Accordingly, the “new type of firm that is emerging” (Zingales, 2000: 1625) is a human capital-intensive organization (Zingales, 2000: 1642ff; Rajan and Zingales, 2000: 214ff; 2001a: 206ff; 2004: 82ff) that needs to elicit the cooperation of managers and employees with rent-creating skills, ideas or strategies who “have been unshackled by the competitive market” (Rajan and Zingales, 2000: 214). Significantly, in these firms, nonhuman assets have become more commodity-like, and can be easily replicated by talented employees literally setting up as competition to exploit growth opportunities that used to “belong” to the firm (Rajan and Zingales, 2000: 214; 2001a: 207; 2004: 84). This possibility alone, according to Zingales

---

108 According to Rajan and Zingales (2001a: 208; 2004: 80; see also Zingales, 2000: 1637), empirical support for this observation is provided by Bhidé’s (2000: 54ff) finding that 71% of the young and fast-growing firms included in the Inc 500 ranking were founded by people who exploited a growth opportunity by replicating or modifying an idea...
indicates that the employer’s authority is “severely limited by the ability of employees to quit, taking with them their human capital or part of the firm.” This means that the locus of power is shifting from owners to employees, and that “the power relationship is not all one-way” any more (Rajan and Zingales, 2004: 83).

The new distribution of outside options weakens the relevance of Hart’s story, and brings Alchian and Demsetz’s observation that it is difficult to distinguish intra-firm authority from ordinary market contracting back to the center stage. Rajan and Zingales (2001b: 840) nevertheless maintained that Coase’s depiction of the firm as “an entity where transactions are driven more by authority or power than by prices” is correct. It follows that the theoretical challenge, emphasized in Zingales’s (2000: 1625, 1644) agenda-setting article, “In Search of New Foundations,” is to overcome Alchian and Demsetz’s critique in the human capital-intensive firm: one needs to explain how power over human capital can be acquired, maintained, and enhanced encountered in previous employment.

It follows that, contrary to Williamson’s accent on low-powered incentives within the firm, skilled employees now need powerful incentives to remain within the firm.

The overall empirical effect of these transformations is that the firm’s size is “shrinking” (Rajan and Zingales, 2004: 80ff), and that its hierarchical organization is “flattening” (Rajan and Wulf, 2006a: 759ff). In a related argument, Langlois (2003: 352ff; 2007b: 18ff) has claimed that the “visible hand” of management characteristic of Chandlerian firms is “vanishing” as markets have thickened.

In addition, it would seem that Williamson’s account of forbearance as the source of intra-firm authority no longer holds in the new type of firm.
when nonhuman assets are insignificant. For Rajan and Zingales (1998: 388), this task amounts to developing a “more general theory of power in organizations, and thus, a more general theory of the firm.”

Generally speaking, on Rajan and Zingales’s (2000a: 204) view, an individual can “derive power from the valuable resources she brings to the production process (and, hence, the resources she can threaten to withhold),” even if production takes place “in environments where property rights are not well defined, are poorly enforced, or cannot be enforced” (Rajan and Zingales, 1998: 390). The human capital intensive-firm, in which “de jure mechanisms are of little direct use in offering [the employer] residual rights over [employees’] human capital” (Rajan and Zingales, 2000: 214), is therefore a case in point, but the overall thrust of the argument also applies to the informal sector in both developed and developing economies. In all these situations, some sort of “de facto mechanism” (Rajan and Zingales, 2000: 216) must substitute for ownership in providing the appropriate investment incentives.

One way to make progress in this direction is to introduce a variation on the classic distinction between ownership and control. Indeed, contrary to

---

112 For Rajan and Zingales, the importance of this task extends beyond the theory of the firm itself, since a theory of the firm is implicit in any discussion of corporate governance and corporate finance, as well as any policy prescription (see Zingales, 1998: 500; 2000: 1629ff; Rajan and Zingales, 2000: 202; 2001a: 206).

113 As well as to “illegal organizations such as the Mafia” (Rajan and Zingales, 1998a: 390). See Dixit (2004, 2009) for a detailed discussion of de facto mechanisms of contractual enforcement in situations of “lawlessness.”
Hart’s definition of ownership as residual control, Zingales (2000: 1638) argued, it is important to distinguish “the role played by ownership, which confers the right to withdraw a resource after specific investments have been made, from that of control, which regulates ... access ... before specific investments are made” (see also Rajan and Zingales, 2000b: 841). In fact, at the formative stages of a business, an entrepreneur’s control over access to a “unique critical resource such as an idea, good customer relationships, a new tool, or superior management technique” (Rajan and Zingales, 2001b: 805) is not only how de facto power is exercised but also the primary driving force behind organizational design. To see how access is relevant on both these counts, it is worth reconsidering some of the fundamental problems of entrepreneurship.\(^\text{114}\)

For Rajan and Zingales (2001b: 805), a key difficulty of entrepreneurship boils down to a trade-off: “how to enlist the cooperation of [other] agents necessary for production without ceding to them too much of the surplus generated by the enterprise.” On the one hand, maximization of the joint surplus depends on the employees’ “firm-specific specialization” (Rajan and Zingales, 2000: 215) that is possible only through “close proximity or access to the critical resource” (Rajan and Zingales, 2001b: 805).\(^\text{115}\) On the other hand, specialized employees, who now have easy access

\(^{114}\) Rajan and Zingales’s story sits well with the idea, popularized by Rifkin (2000) and others, that the shift from ownership to access is a defining feature of the emerging human capital-intensive, service-oriented economy.

\(^{115}\) As Rajan and Zingales (1998a: 388) explained: “if the critical resource is a machine, access implies the ability to operate the machine; if the resource in an idea, access implies being exposed to the details of the idea; if the resource is a person, access is the
to outside sources of finance, are in a position to literally “expropriate this critical resource and compete against the entrepreneur” (Rajan and Zingales, 2001b: 805). Therefore, the entrepreneur’s success crucially depends on his capacity to minimize the risk of expropriation by providing employees with the right incentives to work within the firm rather than against it, so to speak, in the market.

The risk of effective competition from opportunistic employees depends both on the degree of expropriability of the entrepreneur’s critical resources, and on how much access employees actually have (Rajan and Zingales, 2001b: 805). In this context, it seems optimal for the entrepreneur to grant restricted access when the resources are highly expropriable, and to allow greater access when the possibility of expropriation is more limited.\textsuperscript{116} For Rajan and Zingales (1998: 391), this sort of decision is “the essence of internal organization,” properly defined as “the differential access agents within the firm have to the [critical resources] that compose the core of the firm.” Indeed, the objective of internal organization is to “draw forth specialization” (Rajan and Zingales, 2001b: 811) by designing the right

ability to work closely with the person.”

\textsuperscript{116} This is similar to Teece’s (1986: 287ff; 1998: 67ff) argument that an innovator’s ability to appropriate the benefits of innovation, that competitors and imitators attempt to capture, hinges on the “regime of appropriability” defined by the innovation’s degree of “inherent replicability,” and the efficacy of legal protection. Weak definitions of property rights can make assets more replicable, or “fugitive,” as Arrow (1996: 649) put it. In this spirit, Rajan and Zingales (2001b: 831) conjectured that “the relative size of firms in industries with intangible assets should increase when the efficiency of the judicial system improves.” See Kumar, Rajan and Zingales (2001) for an empirical test of this proposition.
balance of power within the firm, that is, by allowing specialized employees to accumulate power and rents that they would not otherwise have in the market. These considerations lead Rajan and Zingales to identify two different organizational structures that can increase the size of the island of conscious power “by having more and more people specializing” (Zingales, 2000: 1646).\footnote{Although the idea that the core of the firm comprises critical resources is similar to the “resource-based view” developed in strategic management by Wernerfelt (1984), Prahalad and Hamel (1990), and others, these contributions, as Rajan and Zingales (2001b: 842) pointed out, do not explain intra-firm power and internal organization.}

When the opportunity for expropriation is limited, the entrepreneur can design a “vertical hierarchy” based on the partial delegation of control over access to specialized employees, who therefore obtain “positional power” over newer generations of employees, and accordingly more rents (Rajan and Zingales, 2001b: 807). The possibility of moving up the ladder ensures that employees remain dedicated to the entrepreneur. However, this organizational structure, typically “found in physical-capital-intensive industries” (Rajan and Zingales, 2001: 839), is less suited when the critical resources are highly expropriable. Hence in human capital-intensive firms, given that the entrepreneur cannot delegate control over access without seriously exposing himself to risk, Rajan and Zingales’s (2001b: 811) argument is that the appropriate organizational structure is a “horizontal hierarchy.” In flat organizations of this kind, employees’ specific investments and loyalty hinge not on positional power over other employees but on the
entrepreneur’s promise to allocate an ownership stake in the firm to a selected few in the future (Rajan and Zingales, 2001b: 839).\textsuperscript{118}

However, although surplus-sharing schemes such as delayed vesting can clearly serve as part of the “glue binding the organization” (Rajan and Zingales, 2001a: 209), the distribution of ownership rights to specialized employees does not necessarily enhance their incentives. Contrary to Hart’s account, as Rajan and Zingales (1998: 406) argued, “ownership may reduce the incentive to specialize.” Indeed, a “dark side of ownership” (Rajan and Zingales, 1998: 390) is that “an owner has a larger opportunity set, and in a variety of circumstances, can face a greater loss of opportunities from specialization” (Rajan and Zingales, 1998: 407).\textsuperscript{119} Accordingly, Rajan and Zingales (1998: 422) concluded, “if all the parties involved in production (i.e., including the entrepreneur) have to make substantial specific investments over time, it may be optimal for a completely unrelated third party to [absorb] the opportunity losses from specialization.” The introduction of outside ownership, Zingales (2000: 1647) observed, may very

\textsuperscript{118} Since this configuration creates an efficiency-enhancing “rat race” among employees, Rajan and Zingales (1998: 418) argued that they had managed to reconcile Marglin’s view that the entrepreneur guarantees his indispensable role in the production process in order to maximize his share of the pie with the classic Smithian analysis of the division of labor (see also Rajan and Zingales, 2001b: 836, n.21).

\textsuperscript{119} The “adverse effect of ownership” (Rajan and Zingales, 1998: 406ff) is that in some situations owners can earn more by selling their stakes than by investing in the business. A similar critique of Hart’s view that asset ownership always improves incentives can be found in Chiu (1998), De Meza and Lockwood (1998), and Wang and Zhu (2005).
well be “the major service venture capitalists provide emerging firms” that they finance (see also Rajan and Zingales, 2000: 233).\textsuperscript{120}

Indeed, the resulting separation between ownership and control over access allows the entrepreneur to concentrate on the coordination and enhancement of the overall specialization, “so as to build complementarities” that “economically link some person or unit that cannot be owned to the firm” (Zingales, 2000: 1645; see also Rajan and Zingales, 2000: 215; 2001a: 207). Complementarities are the real basis of the glue that keeps the firm together. To the extent that the entrepreneur is successful, Rajan and Zingales (2000: 217) reasoned, “the web of past specific investments that creates complementarities between different agents may itself be what is valuable and worth gaining access to.” It follows that “at some point the critical resource becomes the web of specific investments itself” (Zingales, 2000: 1646), and the firm evolves into “something more than a collection of people and resources” (Rajan and Zingales, 2001b: 811).\textsuperscript{121} This is why the firm is normally “able to produce more than any competing entrepreneur with a similar resource starting from scratch” (Zingales, 2000: 1646). According to Zingales (1998a: 498):

\begin{quote}
The firm [is] a nexus of specific investments: a combination of mutually specialized assets and people
\end{quote}

\textsuperscript{120} This is a form of Hobbesian solution. A related contribution of the venture capitalist is the standardization of the firm’s operations (Rajan, 2012: 1193).

\textsuperscript{121} In other words, over time, the firm develops the “organizational capital” (Zingales, 2000: 1646; Rajan and Zingales, 2001b: 842) emphasized by Klein (1988), and wrongly rejected by Hart.
... that cannot be instantaneously replicated [by the market]. Unlike the property rights view, this definition recognizes that all the parties who are mutually specialized belong to the firm, be they workers, suppliers, or customers.

On this view of “the economic essence of the firm” (Zingales, 1998a: 498; see also Zingales, 2000: 1646; Rajan, Servaes and Zingales; 2000: 39), it is clear that “something is more a part of the enterprise when it has greater complementarities with the rest of the enterprise” (Rajan and Zingales, 2000: 218-219). Importantly, as Rajan and Zingales (2000: 219; 2001b: 842) acknowledged, it is equally clear that this economic definition of the realm of transactions governed by power rather than by prices departs from “the legal definition of the firm” that a firm’s boundaries are determined by the common ownership of assets. For Rajan and Zingales, however, the fact that the entities legally defined as firms may not fit their definition of economic organization is not a weakness of their theory, but points instead to the “inadequacies of the legal definition” (Rajan and Zingales, 2001b: 843). Like Hart’s account, the legal definition is “becoming more and more anachronistic and will someday be abandoned” for a “less clear-cut but more realistic” definition similar to one suggested here (Rajan and Zingales, 2000: 219).

The implicit suggestion that the legal system should catch up with the new business realities, because the assumption that “the legal and the economic entity coincide” (Zingales, 1998a: 498) is no longer true, sheds an interesting light on Rajan and Zingales’s agenda. It is important to realize, as Zingales (1998: 498) put it, that “a corporation, in principle, is just an empty
legal shell,” the value of which is derived from “the claims the legal shell has on an underlying economic entity, which I shall refer to as the firm.” If, as Rajan and Zingales tacitly suspect, the corporate form fails to secure claims over the value created by the underlying human capital-intensive firm, then we should expect these firms to adopt alternative legal forms of business organization in order to internalize this externality. On this interpretation of Rajan and Zingales, the comparative efficiency of alternative governance systems, including considerations of capital structure, depends on their relative success in protecting “the integrity of the firm” (Zingales, 2000: 1645; see also Rajan and Zingales, 2000: 222).
I take the ownership of assets by firms, and the attendant feature that economic contracts are made with firms, not their employees or owners, as one of the most significant and robust empirical regularities to be explained by any theory of the firm. Having said this let me confess … that I am unable myself to offer a well-developed explanation of asset ownership by firms.

Holmstöm (1999: 75)

Rajan and Zingales’s emphasis on the efficiency-enhancing properties of third party ownership is at odds with the theory of the firm narrative, and their contention that the traditional legal definition of firms as distinct ownership units are increasingly outdated does little to solve the issue of the determinants of firm boundaries. On the contrary, their otherwise illuminating discussion of power has produced a rather peculiar variation on
the continuum thesis: since the island of conscious power can potentially encompass many legally distinct ownership units, it is clear that the size of the firm, to paraphrase Cheung, is indeterminate because there is no cut-off point. In other words, the problem of organizational anonymity has reappeared. Not surprisingly, despite the progress made since the beginning of the 1970s, the issue of the nature and boundaries of the firm “remains an empirical as well as theoretical challenge” (Garrouste and Saussier, 2005: 179). A case can be made for the fact that this challenge is intimately related to an issue that is largely neglected in the literature, namely the matter of the identity of the central contractual agent.

Explicitly or implicitly, most economists assume that the central agent is a human being (e.g., “entrepreneur,” “employer,” “owner”) or a group of human beings (e.g., “owners,” “management”). This position makes it difficult to come to terms with one of the most “significant and robust empirical regularities to be explained by any theory of the firm” emphasized by Holmström (1999: 75), namely that “firms tend to own or have control over all the alienable, nonhuman assets that they use,” and that “economic contracts are made with firms, not with their employees and owners.” The argument of this chapter is that Holmström’s puzzle reveals an important blind spot in the literature, that most economists seem to be oblivious to, and that this “nexus paradox” is at least in part responsible for the persistence of definitional difficulties (4.1). Arguably, the paradox will remain unresolved as long as theorists of the firm refuse to acknowledge the role played by the firm’s legal personality (4.2). Although this is precisely Jensen and Meckling’s (1976) definition of the firm, the lasting influence of their hostility to “personification” is behind this state of affairs (4.3).
4.1 The definitional impasse

Since Coase (1937), theorists of the firm have struggled to determine the boundaries of the islands of conscious power to be found in the ocean of unconscious cooperation. As we have seen, Coase’s argument that the distinguishing mark of the firm is the replacement of ordinary market contracting by long-term employment contracts, whereby coordination is achieved by the employer’s authority, was contested by Alchian and Demsetz’s (1972) objection that the employer has no more power over employees than customers have over grocers. Coase’s dichotomy view was further criticized by Richardson (1972), Klein, Crawford and Alchian (1978), and other continuum theorists, on the grounds that it conceals a great variety of interfirm contractual arrangements that are strictly speaking neither firms nor markets. To the continuum theorist, the difference between firms and markets is a difference in degree rather than in kind, and what counts analytically, as Cheung (1983) and Alchian (1984) argued, is the effects of different types of contractual arrangements rather than arbitrary labels.

There have been three major attempts to reclaim firm boundaries. Based on the twin notions of asset specificity and opportunism, Williamson (1979a, 1985a) examined the fundamental transformation of ordinary market contracting into interfirm contractual arrangements that foster the parties’ specific investments while minimizing their vulnerability to hold-up risk, and argued that firms, viewed as unified ownership structures that solve disputes internally, emerge in the last resort when asset specificity levels are particularly high. However, Williamson’s account of the comparative advantages of firms had little to say about the costs of integration, and
Grossman and Hart (1986) suggested that redefining integration as the purchase of residual control rights over nonhuman assets not only solves this problem but also leads to a straightforward formalization of firm boundaries and the associated allocation of power. This approach has been revised by Rajan and Zingales (1998) to make sense of the nature and boundaries of the new human capital-intensive firm.

Overall, members of the asset specificity expedition countered the continuum thesis by arguing that the particular constellation of accumulated specific investments allows one to identify a given firm, and to distinguish it not only from the market but also from other firms. After all, they reasoned, “firm-specific” characteristics are not to be found elsewhere. Nevertheless, asset specificity explanations share a crucial, albeit implicit, feature with theories that seem to invite the continuum thesis by presenting firms as kinds of markets. This is “the idea of the firm as a *nexus of contracts* with a central agent” (Ricketts, 2002: 162, emphasis in original). All contractual theorists of the firm agree that the institutional arrangement between various input owners involving contractual centralization enhances efficiency by reducing the number of contracts necessary for the organization of production, thus reducing the associated transaction costs. Cheung’s (1983: 17) denial that the presence of a central contractual agent helps define the firm is the exception that proves the rule.

At the same time, theorists of the firm part ways when it comes to specifying the classes of agents inside the firm’s boundaries alongside the central agent.\textsuperscript{122} Although he later raised doubts about his earlier discussion,  

\textsuperscript{122} This important difference between various theories of the firm has been generally
Coase clearly included an entrepreneur and his employees within firm boundaries. In Alchian and Demsetz’s account the employer is also the owner, but employees are indistinguishable from other agents with whom the owner contracts. This is essentially Hart’s narrow definition of the firm that leads to the exclusion of all agents other than the owner. On the other hand, Williamson’s legally-grounded argument is that, to the extent that disputes arising in employment are normally resolved within the firm without recourse to litigation or arbitration, employees are clearly within firm boundaries along with owners and equity providers, while suppliers, customers and creditors are not. This contrasts with Rajan and Zingales’s claim that it is more realistic to include around an entrepreneur all agents making complementary specific investments, be they employees, suppliers, customers, or investors, but that ownership should be allocated to a third party outside the firm.

Implicit in this discussion is the desirable match between the economic concept of the firm and the corresponding reality of the firm. Expressing this idea, Coase (1937: 386) argued that the definition of the firm needs to be “realistic in that in that it corresponds to what is meant by the firm in the real

overlooked. As a result, some have argued that different theories of the nature of the firm are like the complementary answers that the blind men of Saxe’s poem provide when asked to define an elephant: each isolates one of several features of a complex phenomenon, the firm (see Foss, Lando and Thomsen, 1999: 645; Gibbons, 2005: 239; Garrouste and Saussier, 2005: 179; 2008: 23). However, contrary to the blind men who each isolate one of several features the same, given phenomenon, there is a clear sense in which different theorists of the firm actually isolate different phenomena. A firm comprising classes of agents X and Y within its boundaries is different from a firm comprising only class X.
world.” Interestingly, Coase (1937: 404) claimed that his economic definition of the firm as an island of conscious power distinct from ordinary market contracting was “realistic” precisely to the extent that it matched the corresponding legal category of “employer and employee” (see also Masten, 1988). Williamson (1979a: 238) similarly suggested that the fiction of contractual discreteness needed to be replaced by more realistic analytical categories, and found these categories in various legal categories of contract. Rajan and Zingales (2000: 219; 2001b: 843) refused to follow Coase’s precept on the grounds that prevailing legal definitions failed to match the “new” nature and boundaries of the firm spawned by the financial, technological, and regulatory revolutions.

Arguably, if there is any truth to this claim, then there is a pressing need to consider the extent to which existing economic or legal notions (or both) must be revised. After all, a conception of the firm logically underlies contiguous fields, including strategic management (Conner and Prahal, 1996: 478), corporate governance (Bolton and Sharfstein, 1998: 96), and corporate law (Armour and Whincop, 2007: 432), and these will presumably be affected by the changing nature of the firm. But the problem is, and has been, that a consensus regarding existing notions of what the firm is, beyond the mere fact of contractual centralization, was never reached. Indeed, even the idea of a central contractual agent is sometimes present only implicitly.

---

123 Mäki (1998: 410) clarified Coase’s position: a definition is “realistic or unrealistic in that it either refers or does not refer to a real entity or a set of such entities” in the socio-economic realm.

124 Rajan and Zingales did not reject the principle behind Coase’s precept.
As a result, Eggertsson’s (1990: 158) observation, that the lack of standardized vocabulary and careful definitions “makes it difficult to see whether we are dealing with overlapping or competing theories,” still rings true today.

The lack of standardized vocabulary and careful definitions is true not only of the “firm” but also of other key concepts, including “markets” and “transaction costs.” As they appear in the literature, all of these terms are highly ambiguous.125 Transaction costs, for instance, were initially quite distinctly associated with market exchange (Coase, 1937: 390; 1960: 15; Malmgren, 1961: 401; Williamson, 1971: 114; McManus, 1975: 335), but as firms came to be viewed as kinds of markets the expression gradually referred to a number of contracting costs, including information costs, agency costs, measurement costs, bargaining costs, motivation costs, coordination costs, and enforcement costs, all of which were said to arise in both firms and markets (Alchian and Demsetz, 1972: 785; Cheung, 1974: 63; Jensen and Meckling, 1976: 308; Barzel, 1977: 292; 1982: 48; Klein, 1983: 373; Milgrom and Roberts, 1992: 29).126 Although Demsetz (1988a: 144ff; 1997: 426) attempted to overcome potential confusion by proposing a distinction

125 Other concepts, such as “authority,” “power,” “ownership,” and “contract” are also ambiguous in the literature, and their use is far from uniform. Holmström and Roberts’s (1998: 75) observation, that “there is still no single coherent theory of ownership,” could apply to authority, power or contract as well.

126 It seems disingenuous, however, to claim that the firm arises because of transaction costs when transaction costs can be found inside the firm. It is important to differentiate between different types of costs in order to make sense of the basic Coasean narrative.
between intra-firm management costs and inter-firm transaction costs, he nonetheless concluded that both types of costs are found in both firms and markets.\footnote{See also Masten, Meehan and Snyder (1991: 6). On the contrary, Hodgson and Knudsen’s (2007a) model retained the distinction between the two types of cost (see also Hodgson, 2010: 301).}

Since one type of contract substitutes for another, as Cheung (1983) had argued, it is hardly surprising that “it is often impossible to separate one type of transaction cost from another” (Cheung, 1998: 515). This observation led Cheung (1992: 51) to redefine transaction costs as “institutional costs,” that is, as “all those costs that cannot be conceived to exist in a Robinson Crusoe (one-man) economy” (see also Cheung, 1982b: 650; 1987: 56). Presumably, since Friday’s arrival on the scene transforms the one-man economy, transaction costs are associated with any human interaction.\footnote{This idea is partly explained by the view that transaction costs and externalities go hand in hand, and that transaction costs should be broadly defined as “resource losses incurred due to imperfect information” (Dahlman, 1979: 148). As Allen (1998: 106) argued, any costs that violate the Coase theorem can be defined as transaction costs.} Thus, since “all organization costs are transaction costs, and vice versa,” as Cheung (1987: 56) explained, transaction costs are pervasive “even in an economy where market transactions are suppressed, as in a communist state” (see also Cheung, 1982a: 35). Viewed this way, it is clear that the concept of transaction costs has been diluted to the point of being a catch-all concept,
and its usefulness in explaining the emergence of the firm in a market economy has become limited. 129

Although the entire edifice of the contractual theory of the firm is built on Williamson’s (1975: 20) premise that “in the beginning there were markets,” the meaning of the term “market” is surprisingly unclear in the literature. 130 Whereas the bulk of the literature on vertical integration focuses only on the intermediate products market, the market and its functioning are often treated as a black box (see Holmström and Roberts, 1998: 77). It is often unclear if by “market” is meant something different from “transactions,” and more than “competition” or “market forces.” 131 Of course, as Tirole (1988b: 12) put it, “there is no simple recipe for defining a market,” but this lack of standardized vocabulary and careful definitions does little to

129 Similar considerations led Goldberg (1985: 398) to point out that the concept of transaction cost “runs the risk of becoming … the all-purpose answer that tells us nothing.” Anticipating these problems in another context, Fischer (1977: 332, n.5) observed that “transaction costs have a well-deserved bad name as a theoretical device … because there is a suspicion that almost anything can be rationalized by invoking suitably-specified transaction costs.” Williamson (1995: 33) himself was aware of the “grave problem with broad, elastic and plausible concepts – of which ‘transaction cost’ is one,” pointing out that “concepts that explain everything explain nothing.”

130 This parallels Coase’s (1988a:7) observation regarding conventional price theory, where “the discussion of the market itself ha[d] entirely disappeared.” Useful discussions of alternative definitions of markets can be found in Hodgson (2001b; 2008a; 2008b).

131 This is certainly the case in the new property rights theory of Grossman, Hart and Moore, as well as in its extension by Rajan and Zingales to the human-capital intensive firm.
advance the definition of the firm and its boundaries. Definitional problems are compounded when terms like “internal markets” (Doeringer and Piore, 1971), “surrogate markets” (Alchian and Demsetz, 1972) or “market hierarchies” (Pitelis, 1991) are used to refer to the firm, and matters are not resolved when terms like “quasi-firms” (Eccles, 1981) refer to “mutants” (Hutter and Teubner, 1993) or “strange forms” (Ménard, 1996) that are somehow “firm-like” (Demsetz, 1988a) but somewhere in between firms and markets.

The view that the difference between firms and markets is a difference in degree rather than in kind underpins the tendency to minimize the importance of careful definitions, and leads to the conclusion that taxonomies are speculative, arbitrary, and ultimately pointless. This is why so many economists have fallen prey to the “continuum fallacy” and have abandoned attempts to define the firm (e.g., Cheung, 1983: 19; 1992: 56; Alchian and Woodward, 1988: 76; Demsetz, 1988a: 155; Masten, 1998: 58-59). In this context, instead of thinking that the firm is somehow a substitute for the market, which would imply a coherent definition of the firm (and, relatedly, of the market), a common strategy has involved explaining away the firm by concentrating on something else. This was Cheung’s (1983: 10; 1992: 63) argument that one type of contract substitutes for another. However, the

---

132 The continuum fallacy, that is a variation on the ancient “sorites paradox,” has the following form: (a) A differs from Z by a continuum of insignificant change; (b) no non-arbitrary line can be drawn; therefore (c1) there is no real difference between A and Z; or (c2) A does not exist. The conclusion that contraries (or opposites) are really the same when they are connected by intermediate small differences is clearly a fallacy. See Damer (2008: 137ff) for a brief overview.
The proposed term “contract” is just as vague as the term “firm” (see Schlicht, 1998: 210-211). Both stand for a category, and like any category they have vague or fuzzy boundaries. Indeed, the term “firm” may be, and probably is to a certain extent, vague.\textsuperscript{133}

In economics as elsewhere, categories are normally defined following the Aristotelian slogan, \textit{definitio fit per genus proximum et differentiam specificam}.\textsuperscript{134} Hence, alongside other questions that it is meant to answer, a theory of the firm is about the common features of all the observed forms of the firm (Gindis, 2009: 27, n.1). Each specific form of the firm (e.g., sole proprietorship, partnership, corporation, cooperative, “capitalist,” “democratic”) shares some features with the other forms but also possesses specific or differentiating features not shared with other forms. An Aristotelian reading implies a distinction between “essential” or necessary features to be found in all instances, and “accidental” or contingent features to be found in specific instances only. Alternatively, in line with Wittgenstein’s (1953: §65-77) notion of “family resemblances,” it may be

\footnote{\textsuperscript{133} Vagueness is different from ambiguity. Ambiguity is a conflict between different meanings of a term used in the same context, while vagueness is a problem of linguistic or conceptual precision. As such, vagueness can be tackled by what Copi and Cohen (1990: 136) call “precising definitions.” Carnap (1947: §2) called “explication” the replacement of a vague pre-analytic concept by a new analytically (more) exact one. Useful discussions of vagueness and ambiguity can be found in Tindale (2007: 57ff) and Damer (2008: 121ff). See Kay (2008) for an account of how ambiguity has profoundly affected the development of the theory of the firm, not always for the best.}

\footnote{\textsuperscript{134} In the words of Aristotle (Metaphysics Z, 12, 1037b29): “there is nothing in the definition except the first-named genus and the differentiae.”}
argued that the preceding strategy is an outdated and naïve view of science, and that attempts to define the “essence” or the necessary features of the firm are hopelessly unachievable. On this view, it is misleading to look for a single distinguishing mark of the firm: all the traits described by various theories may be associated with many firms, but any firm may lack any one of these characteristics.

In this context, it is tempting to argue that the “essence of the firm” is nothing more than a Weberian “ideal type” never encountered in reality, and to infer that the ideal features isolated by one theorist need not coincide with those isolated by another. However, it is important to recognize that the term “ideal type” needs to be used with caution because it seems to encourage a “false notion that we are simply dealing with ideas, and not reality” (Hodgson, 2001a: 329). Furthermore, it is important to understand that the process of abstraction, that is an inevitable part of the procedure of theoretical construction, is precisely the identification of what is essential to,

135 Similar ideas can be found in some contemporary theories of concepts and categorization (e.g., Gert, 1995; Fodor, 1998; Laurence and Margolis, 2003).

136 Hence Alchian and Demsetz (1972: 785) observed that “the term firm as commonly used is so turgid of meaning that we can not hope to explain every entity to which the name is attached in common or even technical literature.”

137 From this perspective, Machlup (1967: 28, emphasis in original) described as “ludicrous the efforts of some writers are to attempt one definition of the firm.” Machlup (1967: 28-29) continued: “I hope 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 … Most of the controversies about the ‘firm’ have been due to misunderstanding about what the other specialist was doing.”
and enduring in, any entity (see Hodgson, 2001a: 287). In the process of abstraction, concrete items are stripped from particularities, specificities and exemplifications: ultimately, a universal or quasi-universal is isolated (see Mäki, 1992: 322; Mäki, 2004: 321). The family resemblance argument simply does not provide sufficient reasons for abandoning attempts to define the firm by isolating the essential from the non-essential.  

Arguably, as Ménard (1995: 176-177) and Hodgson (1998c: 35; 2002: 57) pointed out, if our present definitions do not fit reality, this should be taken as a cue to make our definitions more precise, not for abandoning talk of firms and firm boundaries. Machlup (1967: 28) was right about the importance of being “aware of equivocations,” but he was certainly mistaken to conclude that the presence of competing definitions of the firm means that attempts to provide a definition of the firm are “ludicrous.” It is unproductive to think of definitions along the lines of the proverbial “anything goes.” On the contrary, a case can be made for the fact that we need definitions that “carve reality at its joints” (Hodgson, 2001a: 315) so “it can be understood and explained” (Hodgson, 2002: 42).  

As Khalil (1997: 528) similarly observed: “the world was never neat and this has not, and should not, stop theorists from carving it into categories of actors, structures, artifacts, institutions, and so on.”

---

138 Properly understood, the family resemblance argument only means that proposed essential definitions may turn out to be wrong (O’Neill, 2001: 172; Hodgson, 2013: 14). Fallibilism of this kind is hardly problematic.

139 The expression is Plato’s (*Phaedrus* 265e). Plato compared definitions with a butcher’s job of dividing according to objective (in this case physical) articulation.
Some consider that definitions are “mere typographical conveniences” and “theoretically superfluous” (Whitehead and Russell, 1910: 12). However, as Jensen (1983: 329) pointed out, the success or failure of research efforts crucially depends on the choice of definitions. Definitions are of the utmost importance, since in the most fundamental scientific sense to define is to delimit. Definitions, as Suppe (2000: 76) explained, serve to “fix boundaries of phenomena or the range of applicability of terms or concepts.” Indeed, “economic theories differ from one another in that they carve out the world differently” (Mäki, 2004: 319). It is important to understand that definitions do not only define words: “when the words are used they define the world for us,” Samuels (2001: 92) argued, and this means that definitions “may mislead or incompletely define the world” (see also Samuels, 2007: 166ff). From this perspective, a useful theoretical definition of the firm should state what the firm is, and this implies that it is neither circular nor too broad or too narrow. The importance of the preceding considerations for a theory of the firm is easy to understand.

---

140 A definition of any thing or entity comprises an intensive and an extensive aspect, and this implies talk of boundaries: by defining the attributes that the thing or entity possesses and the objects that fall under this category, one necessarily defines the attributes that other things or entities do not possess. Non-resolved debates about the “intension” or sense of the term “firm,” that is, about the attributes shared by all the things or entities that the term denotes, inevitably lead to debates about the term’s “extension” or reference, that is, about the set of possible things or entities the term can denote. See Copi and Cohen (1990: 141ff).

141 Arguably, a definition should not be expressed in ambiguous, obscure or figurative language. These ideas build Copi and Cohen’s (1990: 137) general discussion of
In Williamson’s discussion of asset specificity, firms are organized with common ownership of assets and unified control when highly specific investments are to be made in relatively frequent and uncertain transaction situations. On this view, then, there are no firms, or should be no firms, without specific assets. This hardly seems to be a position a theorist of the firm should be comfortable with: given that there certainly are firms without specific assets, as Gibbons (2005: 209) pointed out, there can and should be a theory of the firm that is not exclusively based on specific investments. By the same token, if the employment relation is the essential characteristic of the firm, as Coase and others have argued, then single-person operations without employees, partnerships without employees, cooperatives, and so on, would not count as firms. Similarly, single entrepreneurs would not count as firms when the firm is defined as an “organization,” a “team” or a “coalition,” where all these notions imply at least two individuals. Again, these are not things a theorist of the firm should be comfortable with.

Of course, the question of theoretically considering the single-person business as a “unitary firm,” as McNulty (1984: 245) put it, will depend on the question at hand, and this will also depend on the scale of observation. For instance, when the focus is on local competition, it makes sense to include single-person operations because these often compete with more complex forms of the firm in consumer markets (see Orts, 1998: 289), but this justification may not hold when considering international trade flows.142

---

142 Similarly, in some cases, defining the firm as an organization may be necessary, theoretical definitions, and on Hodgson’s (1998c: 23ff) more specific discussion of the “Coasean tangle.”
Nevertheless, the issue of the nature of the firm should be relatively independent of instrumental questions of this kind. As Fourie (1993: 43) argued, “one must accept the one-person firm as a conceptual possibility and indeed as an often encountered empirical reality” (see also Dietrich, 1994: 6; Kay, 2008: 1212).\textsuperscript{143} Thus a conception of the firm that fits this reality is needed.\textsuperscript{144} Hence Demsetz (1988a: 156; 1995: 9; 2011c: S10) maintained that single-person businesses logically cannot be excluded because firms should be defined by the fact of production for others rather than for oneself. Are not “firms” production units, as opposed to consumption units, that are present in the circular flow of income in an economy, regardless of their having employees?

However, not all specialized production units that Demsetz would count as firms would be statistically counted as firms. International conventions such as the United Nations System of National Accounts draw a distinction between a production unit that is “incorporated” and one that is not. The former, regardless of whether it has employees, is “a legal or social entity … recognized by law” as distinct from the person who controls it (United

\textsuperscript{143} Over 74\% of registered businesses have no employees in the United Kingdom (Department of Business Innovation and Skills, 2010). This proportion rises to 78\% in the United States (U.S. Department of Commerce, 2007).

\textsuperscript{144} Given this often encountered reality, as Fourie (1993: 44, emphasis in original) observed, the firm “cannot in general have the supersession of the price mechanism as distinguishing characteristic.”
An “institutional unit” of this kind is “capable, in its own right, of owning assets, incurring liability and engaging in economic activities and in transactions with other entities” (United Nations, 2008: 61). This distinguishes it from other forms of production units that typically belong to the household sector.\footnote{More precisely, a production unit will belong to the household when it is not possible to “separate all assets, including financial assets down to the level of cash, into those that belong to the household in its capacity as a consumer from those belonging to the household in its capacity as a producer” (United Nations, 2008: 83). This view was also adopted in the 1993 International Classification of Status in Employment (see International Labour Organization, 1993: 52-55). Conventions of this kind raise important questions regarding the definition of business enterprise in the so-called “informal economy.”} The emphasis on registration is essential. An individual selling things on street corners, however frequently, does not count as a firm, even if the individual in question has a street tradename. On the other hand, registration or incorporation, along with a tradename and other legal, economic and social identifiers, would make the individual count as a bona fide firm. Accordingly, to the extent that these identifiers are nonhuman assets, the Grossman and Hart (1986) view that firms without nonhuman assets are not identifiable in the world (see Khalil, 1997: 524) contains a grain of truth.\footnote{Nonetheless, the Grossman and Hart (1986: 692) definition of the firm “as being composed of the collection of assets that it owns” is blatanly circular, and needs to be avoided (see Gindis, 2007: 270).}

Hart (2011: 102) recently framed the question of the definition of the firm in the following terms: “is a firm circumscribed by its legal status or by its economic activities?” This formulation is misleading because any firm is
circumscribed by both legal status and economic activities. In fact, most economic activities of the firm are only possible because the firm has a legal status, and this is true even though legal and economic boundaries may not always perfectly match. Given the basic principle in most jurisdictions that “only a legal person has the capacity to contract,” “own property,” and “vindicate [its] ownership rights in court” (Iacobucci and Triantis, 2007: 518), markets are populated with legal persons of various kinds (see Deakin, 2006: 318ff). Although all new institutional economists would agree with Coase (1988a: 10) that markets require for their operation the establishment and enforcement of “legal rules governing the rights and duties of those carrying out transactions,” the role played by law’s provision of separate entity status to an entrepreneur or a coalition of resource owners has been largely overlooked. This blind spot has contributed to the persistence of numerous thorny questions about ownership, boundaries, and the sources of intra-firm authority.

4.2 Missing persons

The problem of the identity of the central contractual agent has not attracted much attention. A plausible explanation is that it is assumed that the “entrepreneur,” the “employer,” the “monitor,” and the “owner,” however ownership is defined, is the one and the same human being. To the extent that the basic narrative inherited from Coase is a stylized rationalization of the emergence of entrepreneurial firms in market economies, this implicit assumption may seem appropriate. The entrepreneur, who owns or controls
access to a set of productive assets or resources, and holds residual claim rights to the net value of the output, is the only common signatory of a set of contracts with employees, suppliers, clients, and creditors. An entrepreneurial firm of this kind is a “single ownership unit,” as Grossman and Hart (1986: 693) put it, in which residual control and residual income rights are in the hands of a single agent who thereby obtains power not to be found in ordinary market contracting over all increasingly specialized agents (that may or may not be employees).

This narrative contrasts with Holmström’s “significant and robust empirical regularity” that implies that the central agent holding title to the assets used in production and entering contractual relations with employees, suppliers, and other agents, is not a flesh-and-blood entrepreneur but the firm itself. However, the precise meaning of this statement remains obscure in Holmström’s (1999: 74) own “subeconomy” view of the managerial firm in which “asset ownership conveys the CEO the power to define the ‘rules of the game,’ that is, the ability to restructure the incentives of those that accept to do business on (or with) the island.” Indeed, Holmström’s firm is akin to Williamson’s “unified governance structure” in that “management is centrally implicated in all contracts” with other constituencies (Williamson, 1985a: 318). Given that managers, including the CEO, are usually themselves employees, this setup is similarly at odds with the observation that economic contracts are made with firms and not with their employees. Furthermore, the matter of asset ownership is not settled since managers, including the CEO, typically do not own the assets used in production.147

---

147 Wernerfelt’s (2002: 473) discussion of why the boss should own the assets used
Although Holmström (1999: 75) is “unable … to offer a well-developed explanation of asset ownership by firms,” from which follows the “attendant feature that … contracts are made with firms,” elements of the requisite explanation are available. The key point overlooked by Holmström is the “foundational legal principle” emphasized by Iacobucci and Triantis (2007: 518), that “only legal persons may own property … and … [have] the capacity to contract.” In other words, Holmström’s puzzle vanishes once the firm is viewed as a “legal person” or a “legal entity” (Riordan, 1990: 94; Kreps, 1990a: 753; Milgrom and Roberts, 1992: 20; Kraakman, 2001: 148; Hodgson, 2002: 53; Spulber, 2009a: 52; Triantis, 2011: 190; Deakin, 2012a: 115; Hansmann, 2013: 893) or a “legal agent” (Putterman, 1993: 260; Kornhauser and MacLeod, 2013: 920) with the capacity to own assets and to contract in its own name.\(^\text{148}\) As Van den Steen (2010: 468) explained, the firm “has all the standing of a real person (e.g., it can own assets and write contracts),” even though management exercises these powers by acting “as the ‘as if’ owner of all assets and as the ‘as if’ party to all contracts, thus claiming the authority that goes with centralized ownership.”\(^\text{149}\)

Clearly, legal personality not only allows each firm to be an efficiency-enhancing “contracting hub” (Spulber, 2009a: 440) at the “center of [a] hub-

\(^{148}\) The legal person attributed to the firm by the legal system is quite literally, rather than just metaphorically, the central contractual agent (Hansmann and Kraakman, 2000a: 391).

\(^{149}\) From the legal point of view, the firm and the employer will normally be the same legal person (see Deakin, 2003: 98).
and-spoke network” (Spulber, 2009b: 618) of contracts with other legal persons, it also helps explain the sources of intra-firm authority. Williamson (1985a: 318) recognized that “all contracts are struck with a legal entity called ‘the firm’,” but discounted the importance of this fact in his later examination of the relative advantages of the firm in dispute-resolution respects. Yet his notion of forbearance makes sense only if firms have legal personality. Indeed, courts typically refuse to rule on internal disputes because, as a legal person, the firm “appears in courts as a single party” (Iacobucci and Triantis, 2007: 524).\textsuperscript{150} The attribution of legal personality to firms has implications for legal standing, namely for the determination of the “class of individuals standing to sue in any given dispute” (Holderness, 1998: 508). Since internal divisions do not have legal personality, cannot sue or be sued, cannot own property, and cannot contract in their own name, they do not have an independent standing in court. (Hodgson, 2002: 46; Iacobucci and Triantis, 2007: 524; Gindis, 2009: 39).\textsuperscript{151}

Legal personality enables firms “to enter binding contracts, to seek court enforcement of those contracts, and to do so in their own name” (Milgrom and Roberts, 1992: 20). This means that there is an important sense in which the firm is distinct from any and all of its human constituents,\textsuperscript{150}

\textsuperscript{150} As Wang and Zhu (2004: 96) explained, “forbearance law [implies] that no third-party enforceable ‘procurement contract’ can be used between an employer and an employee.”

\textsuperscript{151} A further implication is that divisions cannot have debts vis-à-vis external parties, since “debt is a personal obligation that is enforceable only against a legal person because only a legal person can be sued” (Iacobucci and Triantis, 2007: 524). These features distinguish intra-firm divisions from subsidiaries in corporate groups.
contrary to Alchian and Woodward’s (1987: 112) view that the firm is a coalition of owners, and distinct from any and all of its nonhuman assets, contrary to Hart and Moore’s (1990: 1120) claim that the firm is a collection of nonhuman assets. The firm may be created by a coalition of owners and may typically hold title to a collection of nonhuman assets, but the firm is not the same thing as the coalition and cannot be said to be equivalent to a collection of assets.\textsuperscript{152} This holds for all observed types of firm: firms are distinct legal agents in various markets, whether they are entrepreneurial or managerial, capitalist or socialist, nonhuman asset-intensive or human capital-intensive, and regardless of the distribution of types. Legal entity status is common to all legal forms that a firm might take, whether sole proprietorship, partnership, mutual, or corporation (Gindis, 2009: 26-27; Deakin, 2012a: 115; 2012b: 353).\textsuperscript{153}

It is difficult to understand why all types of firms have “enhanced transaction capabilities” (Spulber, 2009a: 64) as compared to bilateral exchange in markets without legal personality (see also Armour, Hansmann and Kraakman, 2009: 6). As separate legal persons that “generally can continue in existence after any of the individuals originally involved in it are long gone” (Milgrom and Roberts, 1992: 331; see also Spulber, 2009b: 319),

\textsuperscript{152} Technically speaking the assets belong to the firm as a legal entity, rather than to any individual owner or to the group of owners taken as a whole.

\textsuperscript{153} Contrary to Van den Steen’s (2010: 468) view that legal personality matters mainly for an analysis of large firms with shareholders, legal personality is equally important for the entrepreneurial firm, although for most practical matters it may seem that “the firm (as a legal person) and the owner (as a physical person) are interchangeable.”
firms are able to create many kinds of durable transactions that individual consumers cannot. Accordingly, as Milgrom and Roberts (1992: 331) observed, “reputations can attach to firms rather than just to individuals,” substantiating Kreps’s (1990b: 116) framework in which reputation can rest “in a wholly intangible entity (the firm), as long as those who make decisions or take actions in the entity’s name have a stake in preserving its reputation” (see also Kreps, 1990a: 766). Paraphrasing Hart (2001: 1714), the firm’s independent identity is an important part of the stickiness allowing the firm’s reputation to be separated from that of its key personnel.

These ideas sit well with Williamson’s (1983a: 519; 1985a: 167) observation that “credible commitments … appear mainly in conjunction with irreversible, specialized investments.” As Demsetz (1997: 429) pointed out, in order for the firm’s commitments to be treated as reliable by various long-term contracting parties the risk of asset disgorgement needs to be drastically diminished if not eliminated altogether. More precisely, according to Demsetz (1997: 429), it is clear that

if each of the owners of a small share of equity could insist on the return of his pro rata share of the firm’s assets, as he can, for example, in dealing with an open-ended mutual fund, the firm could be forced to disgorge assets. This would create a great deal of uncertainty for customers and suppliers if these were to find that long-term relationships with the firm are important, for the overhang of the threat to the firm’s assets make the firm’s ability to honor such relationships problematic … To prevent disgorgement of assets, it is desirable for the
firm itself to own the assets once they are acquired with equity provided by shareholders.\textsuperscript{154}

It is now easier to see why ownership rights to assets are vested in the firm itself, and why contracts are made with firms rather than with their owners or their employees. As Blair (2003: 393; 2004: 45; 2005a: 35) observed, capital needs to be “locked-in” to prevent instability and enhance overall credibility. Indeed, locked-in capital can serve as the “hostage” discussed by Williamson (1983), unilaterally committed to ensure the credibility of all the firm’s future transactions, irrespective of any changes in firm membership. This kind of bonding is a key part of the glue that keeps the firm together, and without which, to paraphrase Hart (1995: 59), the firm would indeed be a flimsy entity, “constantly subject to the possibility of breakup or dissolution.” Without the separate legal entity in which ownership rights to assets are vested, “things fall apart; the centre cannot hold.” to quote Yeats (1934 [1920]: 211) completely out of context.

The legal system offers devices that prevent things from falling apart. Most notably, different types of “asset partitioning” (Hansmann and Kraakman, 2000a: 393ff, 2000b: 810ff) or “entity shielding” (Hansmann, Kraakman and Squire, 2006: 1336ff), similar to the civil law notion of “separate patrimony” (Hansmann and Kraakman, 2000a: 439; 2004: 7; 2008: 5, n.4; Armour, Hansmann and Kraakman, 2009: 6), protect in various degrees the firm’s

\textsuperscript{154} Credibility is further reinforced, according to Demsetz (1995: 51), if “the firm … is granted a life that is in important ways distinct from the lives and desires of those who supply it with capital and other inputs.” This is in fact provided for in business entity laws of most jurisdictions.
assets from the personal creditors of its owners, and establishes an order of priority in bankruptcy. These aspects, as Hansmann (2013: 895) observed, have been generally ignored in the law and economics literature, where the focus has always been on “owner shielding” or shareholder limited liability. Nevertheless, a case can be made for the fact that entity shielding, rather than owner shielding, is the most vital aspect of the firm’s continuity over time. Indeed, as Hansmann, Kraakman and Squire (2006: 1336) argued, whereas there are many viable firms lacking shareholder limited liability, “significant enterprises lacking entity shielding are largely unknown in modern times” (see also Hansmann, 2013: 895-896).

Similar considerations led Barzel (1997: 81) to define the firm as “the set of contracts whose variability is contractually guaranteed by common equity capital,” although he did not mention the firm’s attendant legal entity status. According to Barzel (2003: 57), any transaction including this

155 Lamoreaux’s (1998: 70) argument that there are different degrees of “firmness” anticipated to some extent these ideas.


157 In fact, as Blair (2004: 56, n.15) pointed out, “the idea of corporations as separate legal persons for purposes of holding property preceded the idea of limited liability.” The view that entity shielding guarantees continued existence of the firm despite ownership change contrasts with Eggertsson’s (1990: 183) claim that “it is limited liability that guarantees continued existence of the firm despite ownership changes.”

158 As Barzel and Suen (1997: 2) elaborated, firms are “essentially constellations of equity capital surrounding the contracts which [they] guarantee[.]” Hansmann (2013: 893)
guarantee “occurs within the firm and ... is part of the scope of firm” (Barzel, 2003: 57; see also Barzel, 2005: 26). Thus, contrary to Jensen and Meckling’s (1976: 311) dismissal of the distinction between the inside and outside of the firm, and to Cheung’s (1983: 17) claim that there is no rough-and-ready manner to decide whether a given contractual arrangement is “one or two firms,” the acknowledgement of legal personality helps ascertain firm boundaries in a relatively straightforward way (Hodgson, 2002: 44; Gindis, 2009: 39).159 Both market governance and bilateral or hybrid governance involve at least two legal persons, that is, at least two singular firms. In the franchise case, contrary to Rubin (1978: 232), “since each franchisee invests in his outlet [and] has some guarantee capital” (Barzel and Suen, 1997: 4), each franchisee is unequivocally a separate firm. To claim that this may not correspond to the “economic” definition of the franchised organization does not alter the fact that each franchisee is a separate firm.

similarly argued that, “as a general default rule of law, the contracts entered into by a firm – or by any legal entity, including a natural person – are bonded by the assets owned by that entity, in the sense that a patron with an unsatisfied claim against the firm can seize the firm’s assets to satisfy that claim.” Whether or not asset commitment is especially important as a “solution to multi-party action under uncertainty,” as Grandori (2010: 359) has suggested, asset commitment is an essential element of the viability of the firm.

159 The firm can be properly viewed as a “nexus of internal and external contracts” (Reve, 1990: 137ff). In line with Coase’s (1988c: 39) reminder that the emergence of the firm leads not only to the substitution of “intrafirm for factor-factor transactions,” but also to the substitution of “firm-consumer for factor-consumer transactions,” the firm should be viewed as “both ... a nexus of contracts that [indirectly] link its employees to each other, and ... a whole to its customers, suppliers, and stockholders” (Crémer, 1986: 33). See also Hermalin (1999: 106).
The clear functions of legal personality, together with the fact that one finds clear indications of its importance scattered in the literature, raises an important question: why has legal personality been generally absent from the theory of the firm narrative? In part, the answer is that economists are interested in (economic) “substance” rather than (legal) “form,” and have therefore “downplayed or rejected outright the role of the law in defining the firm” (Masten, 1988: 185). There is indeed a widespread view that statutory legal rules and forms need to pass the market survival test to be of interest to economists. As Deakin (2012b: 344) explained, many economists downplay or ignore “the terms used by the legal system” based on the belief that “legal concepts … are ‘fictions’ which are liable to conceal the true nature of the [economic] forces at work.” What matters, from this perspective, is not what is formally provided by the legislator but rather whatever turns out to best suit the needs of rational contracting parties.\textsuperscript{160} To quote Williamson (1990a: 4), since efficiency is enhanced when “individual parties to an exchange ‘contract out of or away from’ the governance structures of the state by devising private ordering” (see also Williamson, 2002a, 2002b), this is what economists should focus on.\textsuperscript{161}

\textsuperscript{160} This kind of argument leads many economists to consider “law as an epiphenomenon of individual interactions” (Hodgson, 2003: 378). See also Hodgson’s (2009) discussion.

\textsuperscript{161} The focus on “private ordering” and more generally on “order without law,” to use Ellickson’s (1991) expression, is a common theme across the new institutional economics (see Richter, 2008). Work along these lines ranges from “self-enforcement” in market transactions (e.g., Klein and Leffler, 1981; Barzel, 2002) to situations of “lawlessness” where the rule of law is either ineffective or absent (e.g., Dixit, 2004). Rajan and Zingales’s
Yet the inclusion of legal definitions does not mean that these should replace or overrule economic analysis. On the contrary, it only means that economists should not forget that parties’ rights and duties are often prescribed by law in ways that cannot be easily altered by contract, if they can be altered at all (see Franke, 1987: 143).\(^{162}\) This is why Behrens (1985: 62) emphasized the necessity to build alongside an “economic theory of the firm” explaining the existence and nature of the firm a complementary “legal theory of the firm” explaining not only the distribution of rights within the firm but also accounting for the legislator’s provision of certain institutional arrangements. Curiously, despite the tendency to ignore legal concepts and to focus on private contracting and market forces, no properly convincing reason “for economists to relinquish a legally-grounded definition of the firm” has ever been provided, as Hodgson (2002: 38) has argued, particularly by “those that wish to dispense with a legally-oriented definition of the firm” (Hodgson, 2002: 55).

A telling case is Rajan and Zingales’s (2000a: 219) rejection of the legal definition of the firm in favor of a “realistic” definition of economic organization. Their claim that the former is anachronistic is based on a misrepresentation. The first and foremost legal definition of the firm is that “each firm is a legal person” (Iacobucci and Triantis, 2007: 515).\(^{163}\)

\(^{162}\) Coase (1992: 717) likewise argued that “the rights which individuals possess, with their duties and privileges, will be, to a large extent, what the law determines.”

\(^{163}\) Interestingly, in a very recent article, Rajan (2012: 1188) recognized not only that “the firm is a separate legal personality defined and protected by law” but also that “there
Economic complementarities between suppliers and buyers, however strong, do not eliminate the fact that contractual relations between suppliers and buyers imply that both are separate legal persons: the relation involves two separate firms. Economic complementarities of the kind Rajan and Zingales have in mind do not eliminate legal structure but are only possible thanks to this structure (Kraakman, 2001: 148). Zingales (1998: 498) correctly pointed out that the value of the legal entity derives from the claims it has on the underlying economic organization, but failed to see, as Schanze (2006: 73) explained, that the legal entity “‘carries’ the real organizational setup; it is the reference point for the enabling framework expressed economically as a nexus of contracts.”

Strikingly, Rajan and Zingales’s account of the third party owner, and of the resulting separation between ownership and control, reinforces the case for the inclusion of legal personality in the theory of the firm rather the reverse. Following Blair and Stout (1999: 274, n.57), the third party owner should be properly viewed not as an individual (venture capitalist or shareholder) but as the “separate legal entity in which ownership rights over assets used in production are vested” (see also Blair, 1999: 85; 2005b: 611; Blair and Stout, 2006: 492), and in whose name central management acts.

---

164 Schanze (2006: 72) could only lament that “there are a dozen ‘theories of the firm,’ meters of shelf space on the ‘theory of organization,’ and very little on the ‘theory of the juridical person.’”

165 Zingales (1998: 501) explicitly acknowledged this interpretation, while Rajan (2012: 1187-1188) has done so only implicitly. Commenting on Rajan and Zingales’s...
Despite different terminologies and objectives, this view is consistent with a generally unnoticed feature of Demsetz’s (1983) work on ownership and control (see also Demsetz and Lehn, 1985). For Demsetz (1995: 50-51), ownership and control are separated not so much between shareholders and managers, but rather between the firm as an entity owning the assets and both shareholders and managers.\footnote{This sort of separation is essential, as Spulber (2009a: 64ff; 2009b: 303ff) has argued, to an understanding of the nature of all types of firm (see also Hovenkamp, 2009: 386ff). To use Rajan and Zingales’ terminology, this separation protects the integrity of the firm.} This sort of separation is essential, as Spulber (2009a: 64ff; 2009b: 303ff) has argued, to an understanding of the nature of all types of firm (see also Hovenkamp, 2009: 386ff). To use Rajan and Zingales’ terminology, this separation protects the integrity of the firm.

Considerations of this kind shed light on Masten’s (1988: 184) view that “the real issue in contemplating the nature of the firm … is not whether the relationships among members are contractual but whether the firm represents a distinct institution,” beyond the comparison between employment contracts and ordinary market contracts.\footnote{As Spulber (2009a: 67; 2009b: 313) observed, firms are “autonomous players” with objectives of their own: “the firm is an independent decision maker as a result of separation from its owners, employees, and trading partners.” This theory, Holmström (1999: 96) pointed out that “the third party could be a firm,” but failed to make the connection with his own puzzle.} According to Demsetz (1995: 50), this is a “largely unrecognized condition” for the entrenchment of management that seems to lead to the conclusion that “the cost of agency … is borne by the firm” (Demsetz, 1983: 376, n.3).

\footnote{According to Demsetz (1995: 50), this is a “largely unrecognized condition” for the entrenchment of management that seems to lead to the conclusion that “the cost of agency … is borne by the firm” (Demsetz, 1983: 376, n.3).}

\footnote{Accordingly, Khalil’s (1995: 49; 1997: 534) claim that institutional underpinnings cannot be the basis of the distinctiveness (or individuality) of the firm is mistaken once legal personality is acknowledged.}
perspective also gives new meaning to North’s (1990: 4ff; 1994: 361; 2005: 59) classic distinction between institutions, viewed as the “rules of the game,” and organizations, viewed as the “players.”¹⁶⁸ Indeed, North’s original failure to explain how organizations can be players in markets alongside individuals was addressed in his new conceptual framework: “the organization must be a legal person capable of bearing rights and duties, and it must be independent of the identity of its individual members at any given moment” (North, Wallis and Weingast, 2009: 152, emphasis in original).

Overall, the inclusion of legal personality in the theory of the firm developed in Coase’s footsteps is unavoidable. Importantly, legal personality allows one to make sense of the efficiency-enhancing contractual structure that is both a key defining aspect of the firm, and the principal justification for its emergence in market economies. Furthermore, theorists of the firm need to recognize that it is only insofar that firms have legal personality that they can own and pledge assets, contract with one another, merge, and act in other market-like ways, not to mention access the legal system in the case of a dispute. These facts are typical of what theorists of the firm purport to explain, but cannot be fully understood as long as legal persons are missing. In a nutshell: without legal persons, there are contracts without contracting parties, and contractual theories of the firm that do not include legal personality are like Hamlet without the Prince. Although the inclusion of legal personality is essential for analytical progress in the field, this progress has been resisted, as is clear from Hart’s (2011: 109) review of Spulber

¹⁶⁸ North’s definition of organizations as “players” is an abstraction rather than a definition (see Hodgson, 2006: 10).
(2009a), that refers the reader to Jensen and Meckling’s (1976) reminder that firms are not individuals with objectives of their own.  

4.3 The reification illusion

As we have seen, only legal personality can explain Holmström’s (1999) significant empirical regularity that assets typically belong to the firm, and that contracts, guaranteed by the firm’s asset commitments, are therefore made with firms as bona fide “contracting entities” (Milgrom and Roberts, 1992: 331; Hansmann, 2013: 893). As independent legal agents “with the right to enter into binding agreements” (Kornhauser and MacLeod, 2013: 918), firms are distinct from their human constituents at any point in time. Beyond the reluctance to adopt legally-grounded definitions of the firm, the implication that firms are “autonomous players” (Spulber, 2009a: 67; North, Wallis and Weingast, 2009: 152) is perhaps the most important reason for the non-inclusion of legal personality in the theory of the firm. After all, as Jensen and Meckling (1976: 306-307) put it, rather than simply assuming that “firms are important actors” in markets, the purpose of the contractual theory of the firm is to look inside the “black box” by focusing on the “conflicting objectives of the individual participants.”

\[\text{169}\]

\[\text{169}\] Like Hart’s (2011) review of Spulber (2009a), Bates’s (2010) review of North, Wallis and Weingast (2009) similarly neglects to mention legal personality, even though legal personality plays a key explanatory role in both books.
The notion that groups of individuals act or can act like single agents is more often than not viewed with suspicion (e.g., Pejovich, 1990: 54; Furubotn and Richter, 1998: 3). This is why when Jensen and Meckling’s (1976) first defined the firm as a nexus of contracts in their landmark article, “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure,” they stressed that “the personalization of the firm … is seriously misleading. The firm is not an individual” (Jensen and Meckling, 1976: 311, emphasis in original). Although they acknowledged “the artificial construct under the law which allows certain organizations to be treated as individuals” (Jensen and Meckling, 1976: 310, n.12), they argued that this is simply a “legal fiction which serves as a nexus for contractual relationships” (Jensen and Meckling, 1976: 311, emphasis in original; see also Jensen and Meckling, 1978: 36; 1979: 470; 1983: 298; Jensen and Ruback, 1983: 43; Jensen, 2003: 1).

The fact that transaction costs are reduced, they explained, when “individuals and organizations – employees, investors, suppliers, customers – contract with each other in the name of a fictional entity” called “the firm” (Meckling and Jensen, 1983: 9) should not obscure the fact that, stripped to their essentials, firms and similar organizations are “pure conceptual artifacts, even when they are assigned the legal status of individuals” (Meckling, 1976: 548). On this view, firms are not really autonomous actors; they cannot be. Firms simply do not have what it takes to qualify as actors. Firms “cannot really have purposes” (Meckling, 1976: 559) or their own “objective functions” (Jensen and Meckling, 1976: 311). Contrary to individuals, Jensen and Meckling (1994: 9) explained, “organizations and groups of individuals cannot have preferences.” Since they lack both purposes and preferences, it is clear that firms “do not choose in the
conscious and rational sense that we attribute to people” (Jensen, 1983: 327). Only “individuals can act and choose” (Brunner and Meckling, 1977: 81), and only individuals can therefore bear the costs (or benefits) of those actions and choices (Meckling, 1977: 29-30, n.36; Meckling and Jensen, 1983: 10).

The problem with legal personality, therefore, is that the notion seems to invite fallacious anthropomorphic conceptions that lead to mistaken economic analyses, specifically when it comes to cost-benefit analyses. As Posner (1990: 186) observed, the fact that “many lawyers and judges … think that one can speak meaningfully of powerful or wealthy corporations, or of placing taxes on corporations rather than on [natural] persons,” is problematic. This position was recently summed up by Werin (2003: 317):

> The very fact that a company is a legal person with its own name, authorized to enter into legally acknowledged contracts and able to make pronouncements of various kinds exactly like a physical individual, leads many people to think of it as a unified, monolithic subject similar to a human being. But the fundamental difference … is that costs and losses – as well as yields – can ultimately fall on individuals only. So, by necessity, any benefit or burden of a legal person

---

170 In his tribute to Meckling, Alchian (2006 [1983]: 267) agreed that “the firm is not an individual entity with a single objective,” and “does not have goals in the sense individuals do.” It is a mistake to think of the firm as an individual entity, he explained, even when all participants in the firm share the same objective.
is, in the end, a benefit or burden to specific individuals as determined by clauses of the contracts entered into by the company (plus legal rules such as tax rates). There is an unfortunate tendency towards misleading anthropomorphism not only in journalistic and political debate, but in professional texts by economists and lawyers as well. This only blurs the picture.

While “ascribing human characteristics to the [firm] is often a useful linguistic expedient,” as Meckling and Jensen (1983: 10) conceded, one should not be snared by what Klein and Coffee (1988: 107) called the “reification illusion.” Regrettably, this illusion, that has “venerable roots in both law and economics” (Meckling and Jensen, 1983: 10), has obscured the fact that “the ‘behavior’ of the firm is like the behavior of a market; i.e., the outcome of a complex equilibrium process” (Jensen and Meckling, 1976: 311; see also Meckling, 1976: 567; Jensen and Smith, 1985: 93-94; Jensen, 1983: 327; 1998: 1). In line with Alchian and Demsetz’s (1972) view of firms as markets, and with Alchian’s (1984: 46, emphasis in original) reminder that “thinking of firms as fundamental actors conceals the *intra*-firm competition” among individuals, Kreps (1990a: 723) captured Jensen and Meckling’s message: “firms are not entities, things of the rough category of the consumer; instead firms are institutions, in the rough category of the market.”

This distinguishes the contractual theory of the firm from

---

171 Kreps (1990a: 724) explained: “firms are places in which exchanges take place between individuals, exchanges that could take place in markets, but which are more efficiently consummated within a firm.” See also Williamson’s (1979a: 239; 1996a: 378) view of firms, markets and hybrid forms as institutional matrices within which transactions
competitive price theory that “treats the firm as it treats the household” (Demsetz, 1992a: 11), namely as “representative individuals” (see Kirman, 1992).  

Strikingly, Jensen and Meckling (1976: 311) observed, although “we seldom fall into the trap of characterizing the wheat or stock market as an individual, we often make this error … about organizations” such as firms. Commonplace errors of this kind give rise to an acute problem: they lead social critics and policy-makers alike to holding firms responsible for things that firms as such cannot commit. Firms cannot “misbehave” any more that they can be said to “behave” in any particular way. Arguably, firms cannot be socially (or otherwise) responsible, since responsibility is an attribute of human beings only. In fact, firms “can no more be responsible than can a lump of coal” (Meckling and Jensen, 1983: 10; see also Jensen and Meckling, 1976: 311).

These considerations about responsibility, that seem at odds with the rest of the theory of the firm literature, were particularly meaningful in the context of the mounting social criticism of big business best represented by

---

172 Meckling and Jensen (1983: 10) adopted an instrumentalist position reminiscent of the realism of assumptions debate of the 1940s-1950s by pointing out that, given its higher scale of observation, the usefulness of competitive price theory and the overall explanation of the functioning of the market system may not be seriously impaired if the firm is treated “as if it were a wealth-maximizing individual.” Nevertheless, they argued, this “anthropomorphic practice … has not been an unmixed blessing” since “it has distracted social scientists away from the study of … organizational form[s].”
consumer activists Nader, Green and Seligman’s (1976a) book, *Taming the Giant Corporation.*\(^{173}\) Just like Alchian and Demsetz (1972) had produced a contractual theory of why capital hires labor as a response to the radical challenge, Jensen and Meckling’s theory was a response to the activism of those that Posner (1974a: 341) called the “Nader-type muckrakers.”\(^{174}\) An important part of the identity of the contractual theory of the firm comes from this engagement with what Fischel (1982a: 1261) more elegantly described as the “corporate governance movement,” that in the late 1960s and early 1970s produced in both academic and popular press a host of variously justified calls for “corporate social responsibility,” “economic democracy,” and other “progressive” legal reforms.\(^{175}\)

\(^{173}\) The only other important contribution to the theory of the firm in which the problem of managerial responsibility was explicitly addressed was Arrow’s (1974) *Limits of Organization.* Contrary to Jensen and Meckling, however, Arrow (1974: 67) argued against simply brushing the matter aside, and emphasized the “functional value of holding authority responsible” (Arrow, 1974: 73).

\(^{174}\) There is a sense in which Nader and his followers, who were reformers but not radicals, composed the conservative side of the New Left. As Posner (1974a: 341) pointed out, Nader’s movement shared with the radicals the belief that “big business – the capitalists – control the institutions of … society.”

\(^{175}\) These pleas for “countervailing powers,” to use Galbraith’s (1952) term, stemmed from the midcentury discussions of imperfect competition, separation of ownership and control, and managerial discretion, in which management was portrayed as “self-perpetuating” (Galbraith, 1967: 73ff; Means, 1969: 20ff) and capable of “administering prices” in non-competitive ways (Means, 1962: 78ff; 1972: 292ff). Alongside these critical works, the managerial theory of the firm (Baumol, 1959; Marris, 1964; Williamson, 1964) explored precisely the same themes.
In fact, a national debate on corporate social responsibility ensued from Nader’s (1965) revelations about General Motors’s reluctance to invest in safety measures in some of their vehicles, and from Nader-backed *Campaign GM* in 1970.\(^{176}\) This highly-publicized campaign, and its implicit endorsement by the Securities and Exchange Committee (SEC), led the firm to create a Public Policy Committee, composed of five outside directors, to advise management on matters concerning the public interest (see Schwartz, 1971a: 432ff, 1971b: 764ff). As Schwartz (1971b: 769) commented, the growing social awareness of the late 1960s had “finally hit the corporate structure in th[e] country.” Importantly, the corporate governance movement pushed for regulation in view of accelerating the process. As Arrow (1973: 316-317) observed at the time, “ethical codes” of conduct “will not develop completely without institutional support” in the form of “taxes, regulations, and legal remedies.”\(^{177}\)

Nader and his followers urged the federal government to take over the chartering process traditionally left to states.\(^{178}\) Federal chartering of large

---

\(^{176}\) Short for *Campaign to Make General Motors Responsible*.

\(^{177}\) This is why Berle (1954: 182ff; 1959: 90ff) combined his hope that managers would develop a “corporate conscience” and act in accordance with a “public consensus” with Miller’s (1959) suggestion to constitutionalize corporate law. The objective was to facilitate business’ integration of the “paramount interest of the community” emphasized in the months preceding the New Deal by Berle and Means (1932: 312-313).

corporations, they argued, would allow government to ensure the independence of a majority of board members, and impose various disclosure requirements.\textsuperscript{179} This would help overcome the failure of states to curb managerial discretion given, in Cary’s (1974a: 663) famous words, the “race to the bottom” initiated by Delaware’s highly permissive (i.e., pro-management) 1967 corporate law statutes. Although Cary, a former SEC Chairman, did not endorse Nader’s “politically unrealistic” federal chartering solution (Cary, 1974a: 700), he did suggest that “the time ha[d] come … to consider a Federal Corporate Minimum Standards Act” (Cary, 1974b: 1101; see also Cary, 1974a: 700ff) in order to protect the rights of investors and consumers from unchecked managerial power. Moreover, even Folk (1972: 312ff; 1976: 1080), the former Chairman of the Delaware statute drafting committee, agreed that higher standards of managerial responsibility would be desirable.\textsuperscript{180}

\textsuperscript{179} Although he did not endorse the federal chartering solution, pointing out that law cannot accomplish everything, Stone (1975: 119ff) argued in favor of increasing social control of the corporate decision-making process.

\textsuperscript{180} Similar concerns were raised at the international level in the late 1960s and early 1970s, and strikingly similar reforms were suggested. Ball (1967: 29), for instance, proposed that it was time to create a truly international company law under which the major international corporations would be incorporated, a suggestion that was under consideration in the discussions leading up to the creation of the United Nations Centre on Transnational Corporations in 1973. In a parallel development, the Commission of the European Communities formally proposed to create a supra-national statute for the European company in 1970. See Rubin (1973) for a detailed comparison.
Following the introduction by a Nader supporter of a bill to establish a Federal Corporate Chartering Commission, the Committee on Commerce of the U.S. Senate held hearings on *Corporate Rights and Responsibilities* in 1976. The pro-federal chartering testimonies of Nader, Green and Seligman (1976c), and Schwartz (1976b), justified regulation by reminding the Senators that the corporation, properly understood, is a “creature of the state” since the corporate charter is essentially a “contract between the corporation and the state” (Nader, 1972: 9). In effect, the state “provides the charter to the corporation, exchanging considerable privileges … in return for certain standards that have to be observed” (Nader, Green and Seligman, 1976c: 207; see also Nader, Green and Seligman, 1976a: 71-73). Given the privileges of legal personality, perpetual existence, and limited liability granted to corporations, Nader inferred, the state should be able to regulate its creations by setting standards of corporate behavior, and sanction socially irresponsible activities. The application of this doctrine at the federal level would avoid the race to the bottom that inevitably undermines individual states’ incentives and abilities to regulate their creatures.

The anti-federal chartering testimonies of Manne (1976), Winter (1976), and Hessen (1976a) attacked the foundations upon which Nader’s argument was built. The view “that the State is the creator and the corporation is the recipient” is false, Hessen (1976a: 14) argued, and

---

181 Berle (1954: 104) similarly wrote: “The corporation is … a creature of the state which charters it … Historically there is sound basis for insisting that the corporation has some color of state authority.” This view sat well with the notion that large corporations are state-like “political entities,” the government of which should be a public matter (e.g., Latham, 1959; Miller, 1960; Eells, 1960; Galbraith, 1972; Dahl, 1973).
designed to obscure the fact that “a corporation is created by a voluntary contractual agreement between individuals seeking to promote their own financial self-interest” (Hessen, 1976a: 9). The state is not a party to the contract; “nor does it give life or birth to the corporation,” since “the role of the State is simply to record the formation of every corporation – nothing more” (Hessen, 1976a: 9). All the so-called “privileges” of corporations can be achieved by contractual arrangement, and in this sense corporations are just like ordinary partnerships (Hessen, 1976a: 21ff).

Crucially, given that “the corporation is … simply a label which denotes a group of individuals banded together in a contractual relationship” (Hessen, 1976a: 14; see also Manne, 1976: 236), the corporation cannot be said to have any rights or responsibilities of its own.

---

182 This argument was central in Hessen’s (1976b) debate with Green (1976) in the press, but also in Hessen’s (1979a: 1332ff; 1979b: 16ff) subsequent writings. Likewise, Meckling and Jensen (1983: 10, emphasis in original) wrote: “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.” Interestingly, Meckling and Jensen (1983: 10) cited Chief Justice Marshall’s famous opinion in the Dartmouth College case of 1819, according to which the corporate charter could be construed as a private contract, and that the legislator’s capacity to modify its terms were restricted by the Contract Clause of the U.S. Constitution (see Horwitz, 1977: 112). For a detailed discussion of this view of corporations see Butler and Ribstein (1989: 770ff).
It follows that any regulatory intervention designed to increase the corporation’s accountability to the public only raises costs for the real individuals involved, particularly the residual claimants. Arguments in favor of intervention, Winter (1976: 45ff) explained, misconstrue jurisdictional competition as a race to the bottom when in actual fact state competition for corporate charters benefits investors by reducing the average costs of incorporation.\textsuperscript{183} The observed characteristics of corporate law result from an evolutionary trial-and-error process, and federally-imposed standards can only interfere with this efficiency-enhancing market mechanism.\textsuperscript{184} In particular, as Manne (1976: 226ff) opined, to the extent that Fama’s (1965a, 1965b, 1970) “efficient market hypothesis” is correct, any attempt to impose outside directors representing the public interferes with the working of the “market for corporate control” (see Manne, 1965: 112ff; 1967: 265ff) that ensures the replacement of less productive managerial teams by more efficient ones. These are exactly Jensen and Meckling’s arguments.

Whatever their shortcomings, Jensen and Meckling (1976: 357) pointed out, corporations have “survived the market test against potential

\textsuperscript{183} See also Winter (1977), Hyman (1978) and Fischel (1978). Dodd and Leftwich (1980) provided empirical support for this observation that overturned Cary’s argument by presenting jurisdictional competition as a “climb to the top” (Fischel, 1982b: 920; see also Baysinger and Butler, 1985a: 182ff; 1985b: 446ff; 1985c: 1270ff; Romano, 1985: 229ff). Jurisdictional competition on the supply side and contractual freedom on the demand side form the basis of what Romano (1993) celebrated as “the genius of American corporate law.”

\textsuperscript{184} As Posner (1977: 307) explained, “competition among states to attract corporations for taxing purposes should result in optimal rules of corporate law.”
alternatives” (see also Jensen, 1983: 328; Fama and Jensen, 1983a: 302; Jensen and Meckling, 1992: 271; Meckling, 1994: 16). Indeed, social critics and reformers have failed to realize that the modern private firm is a human invention having evolved to its present form over hundreds of years, during which “individuals have fashioned a complex network of contractual relations to more effectively serve the objectives of the parties to the legal fiction” (Meckling and Jensen, 1983: 9), and that these are best served when the only “social responsibility of business is to increase its profits,” as Friedman (1970) put it. Any regulations that impose additional costs on the human beings involved, Jensen and Meckling (1978: 32) concluded, can only diminish the chances of the corporation’s survival as a viable contractual choice (see also Meckling and Jensen, 1977: 40).

Tullock’s (1969: 290-291) call for a “new theory of corporations” to counter critics of the prevailing corporate system was answered by Jensen and Meckling’s (1976) arguments. In effect, Jensen and Meckling provided a theory of the firm foundation for the rising anti-regulatory corporate law and economics movement developed in Manne’s footsteps, that allowed Fischel (1982a: 1261) to emphasize Nader’s “lack of understanding of the economic … theory of the firm.” Jensen and Meckling provided not only a timely rhetoric but also a rallying label, “nexus of contracts.” It became common


186 Jensen and Meckling’s influence is manifest in Posner and Scott’s (1980) reader on corporate law and economics edited mainly for law professors: the first section on “The
to argue that corporate law is “trivial,” to use Black’s (1990: 551ff) expression, because all or most statutory corporate features are the product of contractual choice and competitive market forces.\(^{187}\) It also became common to expose the idea of corporate social responsibility as hopeless nonsense. The “entity” that is supposed to “responsible,” it was argued, is a fictitious nexus of contracts that is mistakenly reified (e.g., Fischel, 1982a: 1273; Klein, 1982: 1523; Romano, 1984: 992; Baysinger and Butler, 1985b: 448ff; 1985c: 1270ff; Butler, 1989: 100ff; Bainbridge, 1992: 971; Butler and Ribstein, 1988: 617; 1995: 145ff). In the words of Easterbrook and Fischel (1985: 89):

> The liability of “the corporation” is limited by the fact that the corporation is not real. It is no more than a name for a complex set of contracts among managers.

---

\(^{187}\) Jensen and Meckling (1976: 311, n.14) acknowledged the importance of statutory law, but argued that “common law and human ingenuity in devising contracts” (Jensen and Meckling, 1976: 357) are the ultimate explanations of the evolution of alternative organizational forms. A synthesis of this argument is provided by Meckling (1994).
workers, and contributors of capital. It has no existence independent of these relations.\textsuperscript{188}

Symmetrically, this means that all the so-called “corporate” rights can only “apply to individuals [and] not to imaginary ‘entities’” (Hessen, 1979a: 1328; see also Butler and Ribstein, 1995: 143). All rights, as Jensen and Meckling (1978: 36) concurred, particularly property rights, are “human rights” (see also Jensen and Meckling, 1976: 307, n.6; Meckling and Jensen, 1977: 41).\textsuperscript{189} When sets of rights, such as the rights to contract, to property, to sue, and so on, are nominally vested in non-human legal persons, the real referents of these rights can only be the actual human beings involved, that is, the shareholders (Hessen, 1979a: 1336; 1979b: xv). Importantly, Hessen (1976b: 15) continued, the corporate form, or any other legal form of business entity for that matter, produces no new rights, entitlements, duties or responsibilities that the participating individuals do not already have as individuals. An individual’s rights do not change by becoming a member of an organization: an individual does not acquire more rights, and his rights are certainly not reduced.

From this perspective, it follows that criticism of the U.S. Supreme Court’s 1978 ruling in \textit{Bellotti}, that granted corporations “freedom of speech rights” under the First Amendment in the form of campaign finance, is

\textsuperscript{188} This approach leads to “a benign … view of limited liability” (Ribstein, 1991: 82).

\textsuperscript{189} Barzel (1997: 4, n3) similarly argued that “the distinction sometimes made between property rights and human rights is spurious,” because “human rights are simply part of a person’s property rights.” See also Furubotn and Richter (1998: 81), and Allen (1998: 108).
misguided. Those who believe in the existence of a “business-government symbiosis” (e.g., Fusfeld, 1972; Miller, 1975; Lindblom, 1977), and denounce the Court’s extension of a human privilege to fictitious corporate entities that the Bill of Rights was not designed to protect (Miller, 1979: 82ff; 1981: 24ff), have fallen prey to the reification illusion. They fail to see the self-evident fact that the real referents of constitutional rights such as “corporate speech” rights are the human beings involved, ultimately the shareholders, who are exercising their right to use the corporate form as a collective “medium of expression” (O’Kelley, 1979: 1382; Goss, 1979: 856; Pilon, 1979: 1321) in the marketplace for ideas.\(^{190}\) After all, concluded Easterbrook and Fischel (1989: 1426; 1991: 11), there can be no serious argument about the fact that “the ‘personhood’ of a corporation is a matter of convenience rather than reality.”\(^{191}\)

---


\(^{191}\) In 2010 the Supreme Court’s ruling in the controversial *Citizens United* case similarly acknowledged “corporate political speech rights,” spawning precisely the same criticism (e.g., Nader and Weissman, 2010) and exactly the same line of defense (e.g., Ribstein, 2011).
Chapter 5. The corporate personality controversy

It is said by many that the juristic controversy over the nature of corporate personality is dead. If so we have a corpse, and the opportunity to learn from its anatomy.

Hart (1954: 49-50)

Like Hayek’s (1955: 57) rejection of the anthropomorphic fallacy committed when “attributes of ‘personality’ or ‘individuality’ … are ascribed to society,” Jensen and Meckling’s (1976) vigorous charge against the “personalization” of the firm was a warning against the reckless misuses of ordinary language in the social sciences, particularly for what they considered to be disputable political ends.\(^{192}\) Hence Jensen and Meckling

\(^{192}\) Hayek (1973: 27; 1978: 72; 1988: 112ff) repeatedly argued that many of the enduring “collectivist fallacies” in the social sciences are due to anthropomorphic language.
dismissed Nader’s regulatory proposals by accusing him of falling prey to the reification illusion. Yet Nader certainly agreed with his opponents that corporate personality is a misleading fiction (Nader, Green and Seligman, 1976a: 183; 1976c: 207; Nader and Meyer, 1988: 31), and never claimed that corporations are or can be literally responsible entities. He merely argued that they can and should be made accountable for the public’s benefit.

Although this normative debate about the scope of corporate rights and duties is linked to opposite views regarding the public or private origin of the corporate form, the question of the real or fictitious nature of the corporation itself is also clearly raised.

Economists are generally unaware that these questions have preoccupied jurists for centuries, and that the 1970s debate is a partial replica of a forgotten literature discussing the nature and origin of corporate personality. The argument of this chapter is that theorists of the firm can benefit from revisiting this debate. As the great jurist Hart (1954: 49-50; see also Hart, 1983: 36) pointed out, if “the juristic controversy over the nature of corporate personality is dead,” then “we have a corpse, and the opportunity to learn from its anatomy.” In the late nineteenth century, the old view that corporate personality is a state-created fiction, precisely the perspective resurrected by Nader, was supplanted by a private contract approach strikingly similar to Jensen and Meckling’s ideas (5.1). At the same time, in Europe, an alternative view of the corporation as neither contract nor concession dominated debates (5.2), while parallel American discussions of the “corporation problem,” as many referred to it, also made significant advances in corporate theory (5.3). Although the controversy died out by the 1930s, several insights for the theory of the firm can be derived from this forgotten literature.
5.1 From status to contract

In “The Myth of the Corporation as a Creation of the State,” Anderson and Tollison (1983: 116) unequivocally claimed that “the emergence and spread of the corporation did not depend on legislation” but on “the interplay of market forces.” Likewise, Mahoney (2000: 892-893) argued that “the benefits of treating a business as something separate from its owners are so obvious … that it has never required substantial governmental assistance to achieve.” On this view, corporate personality and other features of “corporateness” can be “created by contract, supplemented perhaps by other common law devices such as trusts or agency” (Mahoney, 2000: 873-874). However, an overview of the history of corporate theory reveals that, overwhelmingly, jurists of the past held the exact opposite position, namely that “the corporation is, and must be, the creature of the state,” to use Maitland’s (1900a: xxx) words. This dogma was self-evident at least as far back as the Roman Empire, and was unquestioned by the glossators and commentators of the Late Middle Ages. The era of papal and royal absolutism, the mercantilist period, and the early industrial revolution were similarly dominated by this view of corporations.

Roman civil law recognized that some groups, formed for the pursuit of a common purpose, were distinct from the collection of their members (see Duff, 1938: 33-37). Typically, universitates (from “forming one whole out

193 Following Plutarch, Blackstone’s (1765: 456-457) Commentaries attributed to Numa Pompilius, the legendary second King of Rome, the creation in the late eighth century BCE of separate collective associations for every distinct trade and social activity.
of many individuals”) and collegia (from “being gathered together”) had what today would be called corporate personality: they possessed the legal capacity of asset ownership that could be seized and sold to cover commercial debts, as well as the capacity of suing in their own name. The organization and operation of these corporate forms was practically uncontrolled under the Republic, but the acquisition of a license from the state became mandatory under the Empire, whose rulers became concerned with potential conspiracies devised behind closed doors by groups of private citizens (Berle and Means, 1931: 414; Duff, 1938: 235). Since the licensing system gave the emperors the ability to modify the conditions of the license by threatening to withdraw it altogether the emperors, as Berle (1928: 4) argued, effectively created a “visitorial power” of the state over corporations.

Berman (1977: 899; 1983: 120ff) and other historians have shown that Western legal science was formed with the rediscovery of the Code of Justinian in the late eleventh century, roughly at the time of the revival of long-distance trade in Europe. This means that the basic framework

Vinogradoff (1922: 120ff) reported that corporate units were also recognized by the Athenians, and regulated by Solon’s laws from the late sixth century BCE onwards. The commonplace view is that “the corporation as a separate legal entity is uniquely European” (Harris, 2009: 613), but some have argued that several legal forms exhibiting at least some corporate features could be found in ancient India in the ninth century BCE (Khanna, 2005), as well as in ancient Mesopotamia at the time of the Code of Hammurabi nearly four millennia ago (VerSteeg, 2000). Very little is known about ancient China in this respect.

Radin (1924-1925: 119) wrote: “fiction or not, and whatever it turns out to be, corporate personality in form and substance was a thoroughly Roman concept.”

Kuran (2011: 100ff) has argued that this rediscovery played an important role in
through which medieval civilists viewed the emerging corporate forms, including boroughs, guilds, monasteries, charities, and universities, was laid down centuries earlier in the Digests by Gaius: *collegia voluntaria* cannot have the status of a corporation unless it is conferred on them by law (Canning, 1980: 16; Bolgár, 1982: 80; Avi-Yonah, 2005: 773; Abatino, Dari-Mattiacci and Perotti, 2011: 368). In other words, corporate personality cannot be formed by anyone at will; its creation is restricted to the acquiescence of the sovereign. Medieval canonists likewise justified the unified jurisdiction and state-like authority of the Pope over various organizations of the Christian community (Berman, 1983: 215ff; Greif, 2006b: 309) that from the eleventh century onward came to be viewed as a corporate body (Pollock and Maitland, 1898: 495ff; Gierke, 1900: 22ff; Lewis, 1935: 49; Kantorowicz, 1957: 194ff; Berman, 1978: 596).

Contrary to the Arab world, where the notion of corporate personality was absent because “Islamic law recognized only flesh-and-blood individuals” (Kuran, 2004: 73), the adoption of the corporate form of property in the West eventually allowed the West not only to emerge from the Dark Ages but also catch up and overtake the enlightened and advanced Arab world. See also Kuran (1997: 63; 2004: 73ff; 2005: 789ff).

According to Maitland (1900a: xvii), in the Digests “there is no text which directly calls the *universitas a persona*, and still less any that calls it *persona ficta.*” Gaius’s principle, as Dewey (1926: 667, n.15) explained, merely stated that “being a *universitas* or *collegium* is something dependent upon statutes … and imperial constitutions.”

A fundamental aspect of medieval religion, according to Tierney (1987: 164), was that the church was not perceived as a collection of separate individuals but was seen as a corporate community, sustaining a life of its own, apart from the secular power. More precisely, as Greif (2006a: 253) elaborated, “because the church had an interest in
It is important to point out that despite its Continental flavor, this theory made sense in the Anglo-Norman feudal system, as a by-product of the Norman Conquest. Since virtually all land became the property of William the Conqueror after 1066, any “land that a person was allowed to retain was considered to be a privilege or concession granted by the Crown, in return for which some compensation was demanded” (Hessen, 1979a: 4). In this context, boroughs and townships seeking some degree of political autonomy claimed special privileges, such as the right to collect local taxes, to organize their own courts, and so on, that they were progressively granted (Pollock and Maitland, 1895: 634ff; Maitland, 1898: 18ff; Raymond, 1906: 356ff; Laski, 1917a: 569; Holdsworth, 1922: 140ff; Pirenne, 1937: 52; Wang, 1942: 498; Frug, 1980: 1090ff). Hence Bracton, the most illustrious of thirteenth-century English jurists, qualified the borough as a *universitas* (Pollock and Maitland, 1898: 676). Describing this “transition from community to corporation,” Maitland (1898: 85-86, emphasis in original) wrote: “the town … gets its capital T … The Town is a person, and may be a landowner among landowners, lessor, hirer, creditor, debtor.”

constituting itself as a corporation,” that is, to be viewed as a power separate from the secular powers of emperors, kings and feudal lords, “it promoted a legal scholarship to define and sanction the legal status of corporations.” The corporate theory used in both ecclesiastical and lay contexts was essentially the same: corporations can be created or authorized only by royal charter or papal bull.

Incorporation, however, came with costs, as boroughs and other new corporations were now exposed to various forms of liability (Bouckaert, 1991: 163). This issue was again theoretically dealt with by appealing to the basic principles of Roman law. By analogy to individual responsibility, the classical jurists believed that corporate culpability must be based on the recognition of collective guilt, and this was clearly lacking. There is no corporate mens rea, they reasoned, given the simple fact that corporations cannot intend and act as human beings can. As Pope Innocent IV famously explained in a 1245 decretal, corporations are persona ficta, that is, “names in law and not that of [natural] persons” (cited in Bolgár, 1982: 82; see also Kantorowicz, 1957: 305).¹⁹⁹ Accursius, one of the great thirteenth-century Italian jurists, similarly claimed that “a corporate body is nothing more than the people who are there” (cited in Canning, 1996: 172).²⁰⁰ These formulations set the terms of the debate, and jurists thereafter multiplied references to the persona ficta that could only be created by the constitutive will of the sovereign.²⁰¹

¹⁹⁹ Following Gierke, Innocent IV is widely credited as the father of this expression (see Maitland, 1900a: xix; Koessler, 1949: 436ff).

²⁰⁰ This belief motivated Innocent’s refusal to excommunicate corporations, and his ruling that sanctions can apply only to the guilty officials (Bolgár, 1982: 81ff; Clarke, 2007: 26). In modern language, Innocent argued that courts should hold individuals responsible by “piercing the corporate veil.”

²⁰¹ For instance, in the famous Sutton’s Hospital case, Chief Justice Coke (2003 [1612]: 355) observed that “incorporation cannot be created without the King,” since a corporation “is only in abstracto, and rests only in intendment and consideration of the law” (Coke, 2003 [1612]: 371), taking this to imply that corporations “may not commit
This view dominated the era of political centralization and royal absolutism. The great English, Dutch, and French chartered trading companies (e.g., Muscovy Company, Levant Company, East India Company, Hudson Bay Company, Virginia Company), that flourished after 1600, were publicly granted monopolies of trade and colonial management (Weber, 1927 [1923]: 211; North and Thomas, 1973: 149; Supple, 1977: 440; Braudel, 1981: 443; Micklethwait and Wooldridge, 2003: 17; Hansmann, Kraakman and Squire, 2006: 1376). At the same time, they were organized as joint-stock companies with tradeable shares, the market for which produced the first financial bubbles. Although proponents of the public choice paradigm (Hessen, 1979a: 29-30; Butler, 1986: 171ff; Harris, 1994: 618ff; 2000: 64ff) have challenged the traditional view that the Bubble Act of 1720 was the regulatory reaction of a panic-stricken Parliament to the financial crash designed to prevent further speculation (e.g., Carr, 1905: 172; Maitland, 1911 [1904]: 390; Formoy, 1923: 28; DuBois, 1938: 2ff), the significant result of the Act was that it forbade the creation of joint-stock companies without an act of crown or Parliament.

---

202 The principle of the sovereign’s absolute power was justified by Bodin (1993 [1576]) in France and Hobbes (1996 [1651]) in England, albeit based on very different rationales. For Hobbes (1996 [1651]: 221), absolute power was weakened by the powers wielded by corporations, that he likened to “worms in the entrails” of the commonwealth.

203 The Mississippi Bubble (1719-1720) in France, and the South Sea Bubble (1720) in England.

204 In the first English treatise on corporate law, written in this context, Kyd (1793:
Similar restrictions operated throughout Europe and the New World. In the United States, charters were granted to ventures deemed to be in the public interest, including commercial banking, insurance, education, railroads, canal-building, bridge-building, navigation, water supply, mining, and so on (see Evans, 1948: 10ff). But state grants became strongly linked with legislative favoritism, and “it became a key feature of post-Jacksonian populist reform to slay the dragon of legislative favoritism by making corporate charters available to anyone who asked for one” (Manning, 1977: 7). Accordingly, the main transformation of corporate law throughout the nineteenth century was the gradual liberalization of business regulation via the widespread adoption of general or “free” incorporation laws that made incorporation a simple administrative matter (Hurst, 1956: 86; 1970: 56-57; 1982: 25; Friedman, 1985: 512; Horwitz, 1985: 187; Millon, 1990: 206ff; 41) declared that “to the existence of all corporations, it has long been an established maxim, that the King’s consent is absolutely necessary.” And only just a few years before the Act’s repeal in 1825, Chitty (1820: 122-123) was equally explicit: “the exclusive right of the Crown to institute corporations, and the necessity for its express or implied consent to their existence, is undoubted.”

Hence Kent’s (1827: 215) claim that “the corporation is a franchise” implied, as Maitland (1900a: xxxi) explained, that corporations hold “a portion of the State’s power.” In the first American treatise on corporate law, Angell and Ames (1832: 3, emphasis in original) similarly wrote: “a corporation is … strictly a political institution.” In this context, American courts strictly applied the ultra vires doctrine according to which transactions beyond the charter’s scope were either null and void or illegal (see Horwitz, 1985: 186ff). This is why, in the 1839 Supreme Court case of Bank of Augusta, Chief Justice Taney ruled that transactions across state borders were null and void: corporations, he argued, only existed within the jurisdiction that chartered them.
Any group of corporators, regardless of the object of their venture, obtained the right to corporate status by simple registration, at a standard fee, provided that a certain set of requirements was met.

In the *laissez-faire* context of the Gilded Age, a period that witnessed the emergence of American industrial capitalism, the doctrine that corporations were public “concessions” gave way to the view that corporations were products of private “rights” and “freedom of contract,” in appearance confirming Maine’s (1861: 170, emphasis in original) famous dictum that “the movement of progressive societies has hitherto been a movement from status to contract.” The 1880s, as Horwitz (1985: 204)

206 Early general incorporation laws were adopted in 1809 in Massachusetts, 1811 in New York, 1837 in Connecticut, 1838 in Florida, 1846 in New Jersey, Ohio and Michigan, and 1849 in California. In Europe corresponding laws were passed in 1844 in Britain, 1867 in France, 1870 in Prussia, and 1883 in Italy. By the mid-1880s, general incorporation with shareholder limited liability was the norm in most Western jurisdictions.

207 An implication of the general incorporation laws was the demise of the *ultra vires* doctrine: corporations could be formed without restrictions on the type of activities pursued.

208 Maine’s dictum was not simply a generalization based on legal history but an illustration of his Victorian *laissez-faire* philosophy (Cohen, 1933: 558; VerSteeg, 1989: 671; Woodard, 1991: 226). At the same time, “status” and “contract” are Weberian ideal-types used to distinguish “archaic” and “modern” societies, much in the spirit of Tönnies’s (1887) “community” and “society,” or Durkheim’s (1893) “mechanical solidarity” and “organic solidarity.”
explained, were characterized by “the tendency to reconceptualize the corporation along partnership-contractualist lines.” In precisely the terms used nearly a century later by Jensen and Meckling (1976) or Hessen (1979a), proponents of the so-called “aggregate theory” of the corporation (Blumberg, 1990: 50; Millon, 1990: 213; Phillips, 1992: 438; 1994: 1065; Dunlavy, 2004: 69; Lamoreaux, 2004: 41; Avi-Yonah, 2005: 771), sometimes referred to as the “associational-contract theory” (Hager, 1989: 580) or the “partnership theory” (Mark, 1987: 1461), proposed that it was time to eliminate talk of “artificial persons” supposedly created by the state.209

The new theory of the corporation was introduced in Morawetz’s (1882) A Treatise on the Law of Private Corporations Other Than Charitable.210 For Morawetz (1882: 24), the fact that in popular conceptions and business practices the key features of corporateness, namely being “considered as personified entities, acting as a unit, and in one name,” had been achieved by unincorporated partnerships, proves that “the idea of a corporation does not necessarily imply a grant of corporate power by statute.” Accordingly, as Taylor (1884: 26), another key aggregate theorist, suggested, if substance is to be preferred over form, the “clearer and more serviceable conception [of the] corporation as consisting of the shareholders” should always be preferred to the misleading focus on the fictitious legal person. After all, as Taylor (1884: 14) explained, in terms of “physical


210 Morawetz’s approach to corporations was later described by Seymour (1903: 549) as “the first attempt to put [corporate] law upon a scientific basis.”
existences” there is only “a collection of [natural] persons” (see also Taylor, 1882: 369; Morawetz, 1882: 1; Thompson, 1894: 395; 1895: 3; Trapnell, 1897: 12; Clark, 1897: 9; 1909: 296-297; Clark and Marshall, 1901: 3; Purdy, 1905: 3-4).

Aggregate theorists believed that “a corporation is only a highly developed partnership,” as Cook (1891: 88) put it, and claimed that “a corporation is an association formed by a contract between the members who compose it” (Morawetz, 1882: 420). This mutual consent, they argued, is essential to the charter that, when granted by the legislature and accepted by the grantees, “constitutes a contract between the State and the corporators” (Morawetz, 1882: 420) that cannot later be altered by the state (Taylor, 1884: 325; Thompson, 1895: 55; Clark, 1897: 51; Trapnell, 1897: 13).\textsuperscript{211} The implication of the partnership analogy, that “the rights and duties of an incorporated association,” particularly as related to property, “are in reality the rights and duties of the persons who compose it, and not of an imaginary being” (Morawetz, 1882: 2), soon found an echo in the courts. In the 1882 case of San Mateo Justice Field opined that if courts were to “always look beyond the name of the artificial being to the individuals whom it represents,” they would realize that “to deprive the corporation of its property, or to burden it, is, in fact, to deprive the corporators of their

\textsuperscript{211} The view that the corporate charter is a contract between the corporators and the state (see also Harvey, 1906: 117-118; Wormser, 1914: 107-108) distinguishes the late nineteenth century version of the contractual approach from the modern version of Jensen, Meckling, and Hessen, although the reference to Dartmouth College and the implication that the state cannot unilaterally alter the terms of the charter are common to both approaches.
property or to lessen its value” (quoted by Mark, 1987: 1460; see also Horwitz, 1985: 1354; 1992: 69; Harris, 2006: 1469).

This point of view was perhaps best expressed by Pomeroy, one of the corporate lawyers involved. According to Pomeroy (cited by Horwitz, 1985: 178, emphasis in the original case transcript),

whatever be the legal nature of a corporation as an artificial, metaphysical being, separate and distinct from the individual members, and whatever distinctions the common law makes, in carrying out the technical legal conception, between property of the corporation and that of the individual members, still in applying the fundamental guaranties of the constitution, and in thus protecting rights of property, these metaphysical and technical notions must give way to the reality. The truth cannot be evaded that, for the purpose of protecting rights, the property of all business and trading corporations IS the property of the individual corporators. A State act depriving a business corporation of its property without due process of law, does in fact deprive the individual corporators of their property. In this sense … there is no real distinction between artificial persons or corporations, and natural persons.

The re-description of corporations as private associations of shareholders, in essence very much like partnerships, was accepted by the Supreme Court in its key ruling in the 1886 Santa Clara case that secured corporate property
from state interference under the Equal Protection Clause of the Fourteenth Amendment.\textsuperscript{212} The reason, as Harris (2008: 317) explained, is the same as in the 1978 \textit{Bellotti} case: in the Court’s pro-business view constitutional rights were being granted to shareholders, not to corporate fictions (see also Horwitz, 1985: 176ff; Mark, 1987: 1462ff; Hovenkamp, 1991: 43ff). Although the Court famously and controversially did not wish to hear argument on the question (see Mayer, 1990: 581), the aggregate theory of corporations was not without conceptual problems, and these were immediately pointed out by various critics (see Horwitz, 1985: 206ff).

Among these, Lowell and Lowell (1884: 3-4) warned that anyone “who confounds a corporation with its stockholders, who says that they are the corporation, or that it consists of its members, not only misstates the legal view of the matter, but is in danger of falling into endless confusion and error” (see also Anon, 1885: 114; Williams, 1899: 3). Indeed, if the corporation is an association of shareholders, then the view that corporations

\footnotesize{\textsuperscript{212} For Davis and North (1971: 77), the \textit{Santa Clara} decision was one of the key changes in the institutional environment that led to American economic growth. As Hurst (1970: 65ff; 1982: 29) explained, the decision established the legitimacy of the corporation, thereby not only directing the entrepreneur’s choice of legal form away from the unincorporated partnership but also paving the way for further liberalization of corporate law. By the late 1880s, the restriction on corporate ownership of stock in other corporations was abolished in New Jersey, soon followed by Delaware and New York. Remarkably, this movement led Wilgus (1904a, 1904b, 1905) to call for federal chartering based on arguments later associated with Nader. Like Cary, Wilgus (1904a: 369ff) emphasized the laxity of state laws, an observation famously picked up in the 1933 case of \textit{Liggett} by Justice Brandeis, who described the competition between states as “a race ... not of diligence but of laxity.”}
are owned by shareholders becomes logically problematic. Arguably, either shareholders are the corporation and hence cannot be said to own the corporation, or shareholders own the corporation but are certainly not the corporation. Perhaps more importantly, as Jones (1892: 78-80) observed, aggregate theorists fail to see that

although it may be conceded that the idea of a corporation as a legal entity is “the impalpable and intangible creation of human thought” … further examination of the question will show that it is both confusing and unwise to speak of this idea as “a mere fiction,” or as “a figure of speech” … The invisibility of the corporate entity is no indication that … it is unreal … All rights are invisible but on that account they are no less real. All legal propositions are ideas, but they are not fictions.

5.2 Neither contract nor concession

The developments of corporate law and political philosophy in the latter half of the nineteenth century radically diminished if not eliminated the state’s involvement in the chartering process. In this context, as we have seen, the aggregate theory proposed by Morawetz (1882), Taylor (1884), and others, was essentially a conceptual and normative reaction to the new legislative framework. Hessen (1979a: 22) argued that their “inherence theory” of
corporations, as he approvingly calls it, namely that “men have a natural right to form a corporation by contract for their own benefit,” with the implication that corporate rights and duties are in fact created and sustained entirely by the exercise of individual contractual freedom, “is the only theory of corporations that is faithful to the facts and philosophically consistent with the moral and legal principles of a free society.” At the time of Morawetz and Taylor, however, not all on the Old Continent would have subscribed to this view of social harmony. In Germany, in particular, an alternative legal philosophy dominated debates.\(^{213}\)

The nineteenth century was that of the historical school of law, an intellectual movement that developed in reaction to both the rationalist and universalist project of the French Revolution and the spreading proposals for codifying German law inspired by the 1804 *Code Napoléon*. In the wake of the deliverance from French rule, the school’s founder, Savigny (1815), an ardent opponent of codification, asserted that law is not a product of reason: law develops in a continuous, “organic” process, as do languages, customs or cultures, and its source can thus only be the historically-specific “spirit of the people,” the *Volkgeist* (see Freund, 1890a: 474ff; Montmorency, 1910: 52; Kantorowicz, 1937: 335; Wilhelm, 1968: 21; Stein, 1980: 60; Gale, 1982: 131; Pearson, 1997: 28; Lindenfeld, 1997: 74; Thornhill, 2006: 133; Freeman, 2008: 1077ff).\(^{214}\) Nevertheless, and perhaps somewhat

\(^{213}\) The terms of the particularly intense dispute in Germany influenced the structure of corresponding discussions in France, Italy, and Belgium. See Hallis’s (1930) detailed discussion of the Continental debates.

\(^{214}\) Savigny’s influence on the German historical school of economics developed in
inconsistently, Savigny (1840) turned to Roman law to provide the material needed by the German nation in its quest for political and legislative unification. This led him to profess that corporate personality is a “pure fiction” (cited in Pound, 1959: 228) created by the state, given that human beings possessing volition are the only natural subjects of legal rights (see Hallis, 1930: 5ff).

Savigny was criticized by other members of the historical school who saw in the age-old Germanic tribal customs the real essence of the German nation (see Thornhill, 2006: 207ff). From this perspective, Gierke’s four-volume Das deutsche Genossenschaftsrecht (The German Law of Fellowship), published between 1868 and 1913, led the attack against the prevailing “straight-jacket of received Roman concepts” (Gierke, 1990: 196). Gierke rejected the state-creation theory of corporations not only for its Roman flavor, but also because, like his contemporaries, he believed in


215 The main ideas of Gierke’s theory of corporations were disseminated in the Anglophone world through Maitland’s famous translation of a section of volume 3 (Gierke, 1900). Other partial translations include Barker’s translation of parts of volume 4 (Gierke, 1934), Lewis’s translation of the introduction to volume 1 (Gierke, 1935), Heiman’s translation of several sections of volume 3 (Gierke, 1977), and Black’s translation of key excerpts from volumes 1, 2, and 3 (Gierke, 1990).
the freedom of association. However, Gierke (1934: 46) dismissed the view that a corporation is “nothing more than a network of contractual relationships between its various members,” and the implication that all corporate rights are to be “necessarily reduced to the collective rights of a number of individuals.”

This view, Gierke (1935 [1902]: 140) argued, erroneously assumes that “reality … shows us only single men as subjective, self-contained unities” (see also Gierke, 1934: 96ff), whereas it is possible to “attribute to … the human association full reality and a unitary character” (Gierke, 1935 [1874]: 169, emphasis in original).

The history of legal and political thought, Gierke (1935 [1902]: 144) observed, is full of “theories of the reality of super-individual unities,” all of which ascribe “some sort of reality … to corporate unity.” Although “this does not exclude the possibility of an illusion,” he conceded, “it encourages us to introduce hypothetically the idea of real corporate unity into the problem of legal corporate personality” (Gierke, 1935 [1902]: 144). For Gierke, there is real corporate unity, or “human associational unity” (Gierke, 1935 [1902]: 139), when an association of individuals, by “partially absorbing their individuality, binds a group of individuals together into a new and independent whole” (Gierke, 1935 [1874]: 169), and thereby becomes “a single entity within a plurality” (Gierke, 1990: 243). In effect, Gierke (1990: 242, emphasis in original) explained, when the bond of fellowship is strong,

216 In the words of his translators, Gierke believed that corporations are “more than mere partnerships or simple collections of individual persons” (Barker, 1934: xxiv), and more than a “lifeless nexus of contractual relations between the associated members” (Barker, 1934: lxi). He therefore set out to show that “contract, that greediest of legal categories … cannot painlessly devour the joint-stock company” (Maitland, 1900a: xxiv).
the organized association tends to acquire a “living collective personality” that “exists independently of the separate personality of its members, and remains immutable and unbroken even when members change” (Gierke, 1990: 242, emphasis in original).

Crucially, since the bond of fellowship can turn a group of human beings into a “physico-spiritual living unity” (Gierke, 1935 [1902]: 152) that is in many ways similar to that of a singular human being, there is good reason to argue that the group, too, “has a will and can translate its will into action” (Gierke, 1935 [1902]: 144).\footnote{This is a variation on Plato’s “city-soul analogy” (see Republic 368e-369a) that has strongly influenced legal and political philosophy ever since. Relevant examples of the view that the state is a real person with a collective will can be found in the writings of natural law theorists (including Althusius and Pufendorf), social contract theorists (such as Reid or Rousseau), and idealist philosophers in both Germany (particularly Fichte and Hegel) and Britain (most notably Gierke’s contemporaries, Bradley and Bosanquet). See Coker (1910) for a detailed overview of nineteenth-century organicist theories of the state, including Gierke’s own position (see Coker, 1910: 76ff). Some, but not all, of these theories were the product of nineteenth century romanticism, the first quality of which being perhaps the tendency to endow inanimate things with life (see Wolff, 1938: 502).} Capturing these considerations in his introduction to Gierke, Maitland (1900a: xxvi) wrote of the corporation that it is

no fiction, no symbol, no piece of the State’s machinery, no collective name for individuals, but a living organism and a real person, with body and members and a will of its own. Itself can will, itself can act; it wills and acts by the men who are its organs as a man wills...
and acts by brain, mouth and hand. It is not a fictitious person ... It is a group-person, and its will is a group-will.

It is important to notice that Gierke’s organicist and vitalist position, that corporations are “real persons” with real volition (see Hallis, 1930: 141ff; Barker, 1934: xxxii; Lewis, 1935: 56; Heiman, 1977: 37), allowed him to use Savigny’s “will theory” of rights, namely the Kantian idea that rights result from the right-holder’s will, in order to include rather than exclude corporations from the realm of law’s actual or potential subjects of rights (see Gindis, 2009: 32). Writing at a time when new types of corporate forms, including trade unions, cooperatives, mutuals, clubs, associations, and political parties were proliferating (see Turner, 1992: 179), Gierke provided a rationale for their gradual recognition as legal persons in both courts and legislation by claiming that “human associations are real unities which receive through legal recognition of their personality only what corresponds to their real [pre-legal] nature” (Gierke, 1935 [1902]: 143). Thus, when the law treats corporations and similar fellowships as persons it is not creating a convenient but nonetheless “imaginary unity” (Gierke, 1935 [1902]: 140).

---

218 Like Savigny, Gierke believed that “in the realm of law, will is the soul of personality” (cited in Heiman, 1977: 38), but turned this claim against Savigny’s fiction theory of corporations along the lines introduced by Beseler (1843), Gierke’s mentor (see Friedrich, 1931: 655; Mogi, 1932: 13; Lewis, 1935: 18; 1968: 178; Dreyer, 1993: 14; Backhaus, 2005: 519; Harris, 2006: 1427).
Indeed, it is “not disregarding reality, but giving reality more adequate expression” (Gierke, 1935 [1902]: 143).²¹⁹

Legal recognition of pre-existing reality plays an important role. As Black (1990: xvii) explained, Gierke thought that “the recognition of real group personality … is essential to human liberty, and that the arbitrary treatment of associations is the hallmark of tyranny.”²²⁰ This view was shared by Maitland, who contributed to the circulation of Gierkean ideas at the turn of the twentieth century not only through the translation but through his own later work as well (see Fisher, 1910: 157ff; Stoljar, 1958: 21; Runciman, 1997: 89ff; Runciman and Ryan, 2003: xi). ²²¹ Mack (1952: 249), Derham

²¹⁹ Proponents of the historical school examined the pre-legal sources of law, and saw that key legal concepts and rules had often codified pre-existing social-economic practices. If this observation was to have any normative implication, then surely it must be that the legislator should attempt to match legal form to historically-specific socio-economic substance rather than impose universal or imported rules. Gierke represented this position as he successfully campaigned to increase the share of Germanic influences in the various drafts of the German Civil Code (see Backhaus, 1998: 9; Harris, 2006: 1439). Gierke was particularly critical of the 1888 draft’s neglect to tackle the problem of acquisition and loss of legal personality (see John, 1989: 114), and its subsequent inclusion owes much his efforts.

²²⁰ Black (1990: xv) argued that there are clear parallels with Tocqueville and Mill on the central role of voluntary associations in modern society. Although he was anything but a radical, Gierke found in producers’ cooperatives the truest mark of human associational unity, and therefore believed that they would be the pattern of the future (Black, 1990: xxv-xxvi). Not surprisingly, Gierke regarded one-man companies that lack the associational element with great disfavor (see Wolff, 1938: 502).

²²¹ Gierke’s organicist ideas, although an original import into jurisprudence, were
(1958: 10), and Hart (1983: 37) may be right that Maitland never fully believed in the German jurist’s mystical notion of “physico-spiritual living unity,” but the fact is, that statements such as, “the personality of the corporation ... is in no sense ... artificial or fictitious, but is every whit as real and natural as is the personality of man” (Maitland, 1900b: 335-336; see also Maitland, 1905: 195), inevitably connect him to Gierke. In any case, some but not all aspects of Gierke’s setup figured prominently in the work of Maitland’s followers.  

Standing on Maitland’s shoulders, a new generation of British corporate theorists established an alternative to both Bentham’s (1864: 8) legal positivist claim that “law ... is the will or command of a legislator” (see also Austin, 1832: 17) and Austin’s (1873: 235) contention that corporations are “persons by a figment, and for the sake of brevity in discourse.”  

Certainly not novel in post-Darwinian Britain. Like many other scholars of his time, Maitland was influenced by Spencer’s (1851) views of the “social organism.” However, Maitland (1883) rejected Spencer’s (1879) individualist ethics.  

Maitland was the intellectual father of the British political pluralists (see Webb, 1958: 45ff; Hirst, 1994: 10ff; Eisenberg, 1995: 65; Runciman, 1997: 64ff), a progressive movement that focused on the place of specific corporations such as churches (Figgis, 1914) or trade unions (Cole, 1917, 1920) within the general framework of the sovereign state (Barker, 1915; Laski, 1917b). “All pluralist theory,” Coker (1934: 170) explained, “shows the influence, on the one hand, of earlier sociological and juristic discussions of the state’s relation to economic and professional groups and, on the other hand, of broader ethical and philosophical ideas as to the value of variety and freedom in self-expression.”  

According to Stein (1982/83: 509-510), Austin may have been the first to introduce the term “legal person” into English.
Maitland’s (1900a: xxiv) terminology, Carr’s (1905) treatise, the first to discuss the latest theoretical developments from the Continent, accordingly opposed Gierke’s “realism” (Carr, 1905: 180-185) not only to Savigny’s “fiction” theory (Carr, 1905: 161-163) but also to Austin’s “symbol” view (Carr, 1905: 177-179). Carr (1905: 154) nevertheless criticized Gierke’s dubious “anthropomorphism” for replacing one fiction by another. Although it is true, to quote Dicey (1905: 154), that “whenever men act in concert for a common purpose, they tend to create a body which, from no fiction of law, but from the very nature of things, differs from the individuals of whom it is constituted,” it is equally true, as Brown (1905: 369) observed, that “the group is not an organism.” Nor is it a “real person” in Gierke’s sense.

Arguably, as Salmond (1902: 350-351, emphasis in original) similarly explained, “a group or society of men is a very real thing, but it is only a fictitious person.” The corporation’s pre-legal “personality” is a fiction, not its legal personality. In fact, from Brown’s (1905: 375-376; 1906: 269-270) point of view, the legal personality of corporations and similar groups is as “real” as that of the human individual. Over time, under very different circumstances, both have been recognized as capable of holding rights and duties, that is, allowed to “rank as a unit in the legal scheme” (Carr, 1905: 185) by an increasingly permissive and yet regulative state. Notwithstanding conventional legal jargon, neither is properly speaking “natural,” if by this term is meant “pre-legal.” Moreover, the problem of rights should not be construed exclusively as a matter of volition. Despite there being, as Brown (1905: 374) pointed out, some sort of “psychical reality” arising from “unities of spirit, purpose, interests, and organization,” for in its absence it is difficult to explain how “the group becomes, or tends to become, a unit”
(Brown, 1905: 396), it seems that “no degree of psychological unity” (Brown, 1906: 269) will ever completely satisfy the will theorist of rights.

This is why Salmond (1902: 234ff, 334ff) turned to the “interest theory” of rights that separated rights from volition. According to this theory associated with Jhering (1852), another important figure of the German historical school (see Seagle, 1945: 71ff; Zweigert and Siehr, 1971: 216ff; Stein, 1980: 65ff; Pearson, 1997: 29-30; Elders, 2005: 568ff; Thornhill, 2006: 197ff; Duxbury, 2007: 25ff), rights tend to be assigned when and where there are interests sufficiently important to be legally protected (see Gindis, 2009: 32). Given that the attribution of rights is typically preceded by a “struggle for law” (Jhering, 1879 [1872]: 64) among individuals and groups within society, we only have rights if “there is something for us, and the power of the State recognizes this and protects us” (Jhering, 1913 [1877]: 49-50, emphasis in original). Significantly, Jhering accepted that the purpose of law is to recognize and protect not only individual but also collective interests, and encouraged their advancement (see Geldart, 1913: l; Wieacker, 1995: 460; Tamanaha, 2006: 3).  

224 The argument that law is both the result of conscious endeavors to solve the complex problems of social existence and a method of reconciling conflicting interests is Jhering’s “epoch-making making” contribution, as Pound (1911: 140) described it, to the sociology of law (see also Gurvitch, 1947: 77; Tamanaha, 2006: 61). Gierke was also seen as a predecessor by Ehrlich (1936: 24) and Gurvitch (1947: 72ff). Interestingly, Gierke had a contrasting influence on the founding fathers of modern sociology. Durkheim’s (1893) views of professional groups and of social reality sui generis have a distinct Giekrean flavor, and this is explicitly acknowledged in the preface to the second edition of *Division of Labour in Society* (Durkheim, 1902: xv), while Weber, a jurist by training having
Although Jhering himself subscribed to a view that Maitland (1900a: xxiv) called the “bracket theory” of corporations (see also Carr, 1905: 178; Pound, 1959: 250), namely that rights ultimately belong to human beings and are vested in corporations for convenience only, the interest theory of rights proved to be useful for corporate theorists who wanted to untie the Gordian knot formed by Gierke’s link between pre-legal group personality and legal personality. Indeed, in the aftermath of Salomon, a landmark case argued in 1897 in which the House of Lords clearly upheld the principle of a company’s separate legal personality, even in the case of a one-man company, Jhering’s approach allowed the pragmatic English jurist, reputed to have “never taken dogmatic theories of any kind much to heart” (Pollock, 1911a: 219), to come to terms with evolving legal conceptions.\textsuperscript{225} From a practical point of view, the valuable result of efforts to move corporate theory in a new direction was that “business men and owners of property would have far less difficulty in understanding their rights” (Hogg, 1906: 177).

A “working theory of the corporation,” as Geldart (1911: 94) put it, should be able to “explain the facts without assuming either a fiction or a real personality” (Geldart, 1911: 97). From this perspective, he continued, the

\textsuperscript{225} See Grantham and Rickett’s (1998) detailed discussion of Salomon.
contractual theory of corporations will be of little service.\textsuperscript{226} Taken literally, the contractual view implies not only that there is a “contract made by every shareholder with every other” but also that there are “contracts between each shareholder and the directors, between each shareholder and every [supplier],” that is to say, that there is a set of “innumerable contracts to which [each shareholder] has not given a moment’s thought” (Geldart, 1911: 97-98). Clearly, Geldart (1911: 98) concluded, “to escape from the fictitious person we have fallen into the arms of the fictitious contract.” Echoing this opinion, Laski (1916a: 420) similarly observed that “the contractual theory … can result in fictions compared to which the supposed fiction of corporate personality has less than the ingenuity of childish invention.”

For Laski (1916a: 421), the issue of “property rights serve[s] to bring out the failure of the contractual attitude with striking clearness.” The claim that the property of the corporation is the property of its corporators because the corporators pooled their resources on a voluntary basis is simply contrary to the facts. An even cursory examination of the “true legal nature of corporate ownership” (Smith, 1914a: 69-70; 1914b: 143) reveals that once incorporated, the property thus pooled together belongs to the corporation itself as a matter of law. By the same token, the view that all the rights attributed by law to corporations belong to its human members is equally problematic (see Brown, 1905: 367). It is in their own interest that individuals sometimes give up some of their property rights because this allows them to realize the benefits of collective action, and the primary legal

\textsuperscript{226} As Brown (1905: 367) put it, “the corporation is no mere partnership … Whilst partnership can be resolved into mere contract, a corporation implies a new status.”
device for achieving these benefits is by incorporation, that is, by filing for and obtaining a legally endowed collective capacity for property, contract and litigation (Laski, 1916a: 424; see also Vinogradoff, 1924a: 595).

Whether or not this capacity is granted, and under what conditions, is ultimately a question of public policy (Smith, 1914a: 71; 1914b: 145). Political pluralists generally agreed with Vinogradoff (1924a: 604) that “the life of groups has two sides, i.e., the social contents which are real and produce the union, and the legal form which has to be arranged artificially by the State in order to safeguard public and private interests” (see also Vinogradoff, 1924b: 550). It follows, they reasoned, that it is a matter of policy to decide whether corporations should be held responsible for the actions of their directors or managers, and that this could be settled without focusing on the presence or absence of the corporation’s mens rea. Indeed, as Laski (1916a: 409; 1916b: 122ff) observed, given the well-established principle of vicarious liability the corporation can certainly be held responsible for torts, but also for non-compliance with labor laws, and so on (see also Geldart, 1911: 102; Baty, 1916: 65ff; Vinogradoff, 1924a: 602).

227 These questions were central in the debates surrounding the 1901 case of Taft Vale, a case in which the House of Lords ruled that trade unions could be held liable for damages caused by their members. The unions viewed Taft Vale as a barrier to the right to strike, and one of the first significant political actions of the newly-formed Labour Party that entered Parliament in 1906 was to successfully promote the Trade Disputes Act of 1906 that effectively overturned the decision (see Martin, 1958: 118ff). A detailed account of these events is provided by Geldart (1906: 206ff). See Webb and Webb (1920: 596ff) for a general overview.
The state, from this perspective, can regulate corporations even though it does not create them.

5.3 The value of corporate personality

In the United States, where the movement from status to contract seemed to have reached an advanced stage, at least in comparison to Europe, the aggregate theory of corporations turned out to be surprisingly short-lived. Paradoxically, although Santa Clara and the legislative ethos of the 1880s seemed to have legitimized the corporation as a private institution by using the partnership analogy, thereby favoring the emergence of big business, it was precisely the emergence of big business that led to the demise of aggregate theory by the end of the 1890s. As Horwitz (1985: 176; 1992: 74) and many others have explained,\textsuperscript{228} it became difficult to argue that corporations, particularly the increasingly large-scale, concentrated manufacturing conglomerates, were essentially like partnerships.\textsuperscript{229} Indeed, it


\textsuperscript{229} Although at the dawn of the twentieth century only about 14\% of manufacturing establishments were operated by corporations (Evans, 1952: 486), the wave of horizontal integration in American industry that began in the mid-1890s after the passage of general incorporation laws for holding companies in New Jersey led to extreme concentration. According to Lamoreaux (1985: 2-3), more than half of these “consolidations” absorbed
became hard to deny the reality of corporate organization, as something distinct from the aggregate of its shareholders, when vast economic power lay in the hands of professional managers and “absentee ownership,” to use Veblen’s (1923) celebrated expression, was more often than not the norm.

In the years preceding Veblen’s (1904) caustic critique of the prevalence of pecuniary business interests over industrial employment, the “corporation problem,” as Cook (1891) referred to it, namely the pressing societal need to deal with the numerous abuses perpetrated by corporate promoters and managers exploiting their oligopolistic positions, had been at the center of public, academic and legislative debates (see Hurst, 1970: 61ff). Like many of their contemporaries, progressive economists such as Ely (1887: 977) observed that an important part of the difficulty of finding “some contrivance which will render artificial persons amenable to the moral law” of liability was due to the fact that “the nature of corporations ha[d] not yet been fully explained” (Ely, 1887: 975).

The need for a new theory of over 40% of their industries, and about a third absorbed over 70%. Chandler (1990: 75) described these events as “the largest and certainly the most significant merger movement in American history.”

Congress created a federal Interstate Commerce Commission in 1887 to tackle the problem of pricing abuses by railroad corporations (see Sklar, 1988: 51; Berk, 1997: 100ff), and took steps to restrict the widespread use of the trust arrangement as a means of avoiding corporate law regulations with the Sherman Antitrust Act in 1890 (see Hovenkamp, 1991: 64; Friedman, 2002: 53ff).

This view was shared by Adams (1891: 75ff; 1897: 16ff) and other precursors of American institutionalism, including sociologists such as Small (1896: 403). In general, the issue of the “social control of business” (Clark, 1926) was central for members of the
corporations was particularly manifest in the debate surrounding the problem of corporate taxation. The Revenue Act of 1894 included a first attempt to tax corporations at the federal level (see Joseph, 2004: 76ff; Bank, 2010: 40ff), and this required some justification.\textsuperscript{232}

From an analytical point of view, extant corporate theories seemed unable to come to terms with what Davis (1897: 274) described as the “new social forces” associated with the rise of the business corporation. In particular, the aggregate theory failed to explain the “voluntary inception” (Davis, 1897: 280) of corporations without reducing the charter to a “contract” or a “bargain between a state and a group of its citizens” (Davis, 1897: 282), an idea that Williams (1899: 68), among others, claimed to be “founded partly upon a misconception of the true nature of corporations, and partly upon the failure to distinguish between ancient and modern charters.” The idea of a bargain with the state is misleading, Williams (1899: 74) argued, because the terms of the bargain cannot be really altered: after incorporation the rights and duties of shareholders are “fixed and determined … by law, the fact being that such rights and duties cannot in any way be affected by any contract among themselves” (Williams, 1899: 74).\textsuperscript{233} In

\begin{footnotesize}
\textsuperscript{232} See Seligman’s (1890) revealing discussion of the issues, particularly the thorny problem of double taxation. Although “the nature of the corporation [was] still under discussion” (Pepper, 1895: 296), the Act was struck down by the Supreme Court in 1895, only to later uphold the Corporate Tax Act of 1909 (see Bank, 2010: 66ff). The principle of corporate taxation was firmly established in the Revenue Act of 1913. An overview of the ensuing debates is provided in Ballantine (1921).

\textsuperscript{233} This led Williams (1899: 74) to observe that the position of a shareholder in
addition, he continued, although it is true that, unlike the privilege-conferring charters of old, the new charters are merely records of privately initiated incorporations, in both cases the practical result is the provision of artificial personality by the state, and this truth could not be evaded.

To say that a corporation is an artificial person, observed Williams (1899: 10), “is not to deal in metaphysics or in subtle concepts, but is simply to state a legal fact of the greatest practical importance.” Unfortunately, as Davis (1897: 275, n.1) could only lament, the notion of artificial personality has been the “source of much confusion in legislation and legal decisions relating to corporations,” the main difficulty having stemmed from “the effort to apply to them legal principles elaborated in a system of law founded on individual social units instead of modifying the existing system so as to make its principles applicable to [composite] social units.” From this point of view, Davis (1897: 286-287) argued, the fact that an incorporated group acts and is acted upon as a composite unit needs to be acknowledged, and a link between this “compulsory unity” and the notion of artificial personality needs to be established. In this context, Freund’s (1897) The Legal Nature of Corporations represented an important innovation. Significantly, in an effort to meet the “demands of technical jurisprudence” (Freund, 1897: 83), Freund offered the first American discussion of Continental debates.

relation to the corporation is “one of status and not of contract.”

234 Given that “the members of a corporation act not as units, but as parts of a composite unit,” Davis (1897: 279, n.1) explained, “their social relations are ... exceptional as compared with the regular social relations of individuals regarded as social units.”

235 Freund encountered Gierke’s theory before Maitland’s translation (Gierke, 1900),
Freund (1897: 49) attributed the failure of “prevailing theories of corporate existence” to the “orthodox view” of the nature of rights, namely that “undivided personal volition [i]s essential to the holding of a right.” The will theory of rights, Freund pointed out, is the erroneous belief that leads Morawetz, on the one hand, to reduce all corporate rights and duties to the rights and duties of their individual members (Freund, 1897: 12), and Gierke, on the other hand, to see “every corporate act [a]s the manifestation of corporate will and therefore of corporate personality” (Freund, 1897: 76). For Freund, a more adequate theory of corporations is one that rejects Gierke’s “strained view of the corporation as a real person” (Freund, 1897: 76) or “a new and distinct species of humanity” (Freund, 1897: 52) while accounting for the undeniable fact that for most if not all practical purposes law treats “corporation and member as two absolutely different [rights-and-duties] holders” (Freund, 1897: 41). The challenge is to show that corporate rights and duties are not “abnormal and illogical” (Freund, 1897: 48), and that “the treatment of the corporation as a distinctive legal person” (Freund, 1897: 83) is justified.

From Freund’s perspective, the fact that this “technical conception … has grown up in connection with property and not with governmental rights” (Freund, 1897: 9) becomes clearer once rights are construed, following

having studied in Germany between 1881 and 1884 (Kraines, 1974: 2). He not only transmitted German ideas but also “adapted them to the American reality” (Harris, 2006: 1435; see also Horwitz, 1985: 218). In many ways, Freund had the mind of an economist, and his book could arguably be described as the earliest “rational study” of corporate law, to paraphrase Holmes (1897: 1001). Morawetz’s contribution was nowhere nearly as scientific or systematic as Freund’s, contrary to Seymour (1903: 549).
Jhering, as “legally protected interests,” the benefits of which are secured by some “active … element of control” (Freund, 1897: 15). The advantage of identifying the “two elements of a right, control and interest” (Freund, 1897: 17), is that both their normal coincidence in the same natural person and their frequent separation can be explained within the same framework. The latter case is of particular interest to the corporate theorist, as corporations are certainly characterized by a “separation of control and interest” (Freund, 1897: 16). Indeed, it is the specific form of separation, and more precisely the specific form of control, that distinguishes corporations from partnerships and other unincorporated associations. All these forms of association should again be distinguished from standard agency or trust arrangements that similarly involve a separation but lack the “voluntary acts of mutual connection” (Freund, 1897: 22) that are essential to associations.

In any association based on the “combination of resources,” Freund (1897: 19) explained, the purpose of which is to “bring returns to each party far in excess of what he would procure by the separate and independent

---

236 Freund hence anticipated Maitland’s followers. Jhering’s (1879 [1872]) jurisprudence was already familiar and admired in America (see Tamanaha, 2006: 61ff; Duxbury, 2007: 28). In The Common Law, Holmes (1881: 208) had referred to him as “a man of genius.”

237 A related case in point identified by Freund (1897: 22) is the set of “relations in which a number of people are subject to a common authority without having joined each other by voluntary acts of mutual connection. In these cases there is strictly speaking no association, but a sum of individual contractual relations entered into by one person with a number of persons acting separately, and affecting them alike. There is an aggregate body … but for legal purposes its existence is irrelevant.”
employment of this own means,” a key issue that arises is whether the associates must act in concert to pursue their common interests. Given that disagreements over the proper course of action may arise, and in fact often do, particularly when the number of joint parties is large, the obstacle of “concurrent action” (Freund, 1897: 23) needs to be overcome. A common solution to this problem is the adoption of a decision-making procedure such as a majority rule (see Freund, 1897: 24ff), and this implies, according to Freund, a form of separation of control and interest: when associates agree to “principle of representative action” (Freund, 1897: 47), each associate accepts the possibility that the representatives of the majority in control of the common interests may initiate actions that they may not individually desire. It follows that the associates’ commitment to the association, despite this possibility, is of central importance. Without it, the unity and the continuity of the association over time are uncertain.

Indicative of the level of commitment is the associates’ choice of legal form. In fact, without legal form, an association will not only be “legally irrelevant” (Freund, 1897: 71) but will also be unable to exercise any “binding power over its members” (Freund, 1897: 21) since the undivided control over the association’s affairs will not be legally secured. An association of this kind will be constantly subject to the possibility of breakup and dissolution, to paraphrase Hart (1995). On the other hand, depending on the chosen legal form, “security both against outsiders and against defection on the part of the members” (Freund, 1897: 22) can be more or less successfully achieved. Freund’s argument is that incident to the adoption of a particular legal form is the legally binding separation between control and interest, and that this institutionalization of the principle of representative action is what provides the associates with the requisite
security. The greater the degree of separation recognized by law or the courts, the greater the security, and the more “the idea of the unity of the association as a holder of rights is justified” (Freund, 1897: 77).

In Freund’s (1897: 24ff) terminology, the operation of the principle of representation that is an incident of, or originates with, the legal form itself is “original representation.” American courts, he observed, effectively recognize a form of original representation in partnerships. Courts have repeatedly upheld the business convention that “each partner is the representative of the firm” (Freund, 1897: 26) based on the reasoning that a partnership may outwardly “appear as a unit” (Freund, 1897: 30), that this “outward unity is expressed by a collective name and title” (Freund, 1897: 40), and that in ordinary business dealings this implies that partnerships contract their own debts secured by the property held in common. With partnerships, however, the recognition of the separation is imperfect. While in some respects “the undivided control of partnership affairs may be strengthened by contractual stipulations,” Freund (1897: 27) clarified, partnerships lack an absolute protection against “express dissent” and “an unqualified recognition of majority rule” (Freund, 1897: 28). It follows that

---

238 Original representation is fundamentally different from “representative action under express delegation, by which joint rights are commonly exercised” based on the law of agency (Freund, 1897: 23). The main problem with “the relation between agent and principal” is that it “does not solve entirely the difficulties of concurrent action” (Freund, 1897: 23-24).

239 See Pepper’s (1898: 299ff) complementary discussion of partnership property. An extensive list of court cases in which partnerships were treated as de facto legal persons in line with business practices can be found in Cowles (1903).
“the principle of original representation fails in transactions beyond the ordinary course of business” (Freund, 1897: 27), where acquired reputations tend to be insufficient.

By contrast, Freund (1897: 81) argued, original representation that comes with corporate status means that “corporations constitute distinctive parties to legal relations” irrespective of the scope of business, reputation or other factors that affect the unincorporated partnership, and ultimately condemn it to relatively small sizes. Indeed, a fundamental advantage of the corporate form is that undivided control over corporate assets lies in the hands of a “governing body,” whose position is “different from that of mere agents” because its binding powers cannot be revoked at any time, even “by majority [shareholder] resolutions” (Freund, 1897: 59). The separation of control and interest is accordingly complete, with the implications that “the rights held by a corporation are not the rights of any physical person, that its members are not the part owners of the corporate property, nor part creditors or debtors of the corporate claims and obligations” (Freund, 1897: 9-10). An

---

240 Specific members of the governing body may be under some conditions replaced but the undivided control over corporate affairs by a governing body is an incident of the corporate form itself. These considerations led Mark (1987: 1474) to describe “Freund’s brilliant exposition of the notion of undivided control over property within a corporation” as the “intellectual foundation of the position of management,” that anticipated Berle and Means’s (1932) classic discussion by a few decades (see also Mark, 2006: 1485). From this perspective, Freund anticipated aspects of both Blair and Stout’s (1999, 2006) “mediating hierarchy” model of corporations, and Bainbridge’s (2002a, 2002b, 2008) “director primacy” model. Relatedly, but perhaps more importantly, Freund’s exposition was the first to demonstrate the functional value of corporate law.
incorporated association becomes a genuine “property-holding body” (Freund, 1897: 9) that can survive changes in membership, and this is what provides the ultimate security both against outsiders and against defection on the part of the members.

The upshot is that the three “salient features of the body corporate,” namely “its unity, its distinctiveness, and its identity in succession” (Freund (1897: 47), justify the treatment of the corporation as a “human agency devoted to distinct purposes” rather than merely as a “number of individuals” (Freund, 1897: 81). That a qualified departure from the principle of individual rights is necessary in order to make sense of the idea of corporate rights and duties can be understood in a straightforward manner by Freund’s (1897: 77-78) “analogy of composite things”:

If we treat a house, a ship, a forest, or a mine, as one thing, we do not deny that this thing is composed of many separate or severable parts, each of which may be a thing by itself. But in so far as the connection is operative, the part has no legal existence except as a part, and does not form an object of separate legal disposition; it shares the legal status of the composite thing, while as soon as the nexus is broken or only disregarded, it becomes a subject of independent treatment in law. In like manner we treat the [corporation] as one, disregarding the separate existence of its members as individuals, in so far as their recognition as such would make the protection of joint
interests an impossibility, i.e., in so far as it would disturb the conditions of undivided control.

When Morawetz and others claim that the corporation is “in strict logic a fiction,” they fail to see that “fictions based upon the neglect of the irrelevant are very different from fictions which mean the substitution of an imaginary conception for a substantial nonentity” (Freund, 1897: 78). The dogmatic belief in the “axiomatic proposition that the aggregate is nothing but the sum of its parts” (Freund, 1897: 11) accordingly leads them to “unwarranted and fallacious conclusions” that any theory of corporate existence, corporate acts or corporate rights and duties, deals with “undemonstrable entities” (Freund, 1897: 52).

This contrasts with Freund’s own approach that proves that it is possible to satisfy the demands of technical jurisprudence “without having recourse to a fictitious entity” (Freund, 1897: 83), or to the more dubious ideas associated with Gierke. At no point does Freund deny that the corporation “becomes visible and active in and through individuals only” (Freund, 1897: 77). On the contrary, Freund’s theory is built on the fact that, far from being speculative, the idea of “corporate acting capacity” (Freund, 1897: 55) is directly derived from acts by individuals in appropriate corporate positions: “when we speak of an act or an attribute as corporate, it is not corporate in the psychologically collective sense, but merely representative, and imputed to the corporation for reasons of policy and convenience” (Freund, 1897: 52). Indeed, Freund (1897: 75) argued, “when

241 See Fuller (1967) for a detailed discussion of different kinds of fictions to be found in jurisprudence and language more generally.
we speak of corporate acts, or corporate tort … we mean … representative acts [or] tort.” As an “instrument of legal reasoning,” the objective of which is to “determine the incidence of the effects of legal acts done in the corporate name” (Freund, 1897: 82), there is nothing objectionable in the “idea of vicarious performance,” Freund (1897: 56) concluded.

Freund’s ideas sat well with the views of some of his contemporaries.²⁴² Among these, Robinson’s (1900: 80) assertion that every “voluntary combination” formed for the pursuit of common interests is necessarily “unstable” unless “its existence is rendered imperishable by law,” seems to be an echo of Freund’s account of the value of corporate personality.²⁴³ Other prominent jurists, such as Pepper (1901: 267), explicitly emphasized the superiority of Freund’s representation theory over the Gierke’s organicist approach. Deiser (1909: 314) likewise acknowledged that Freund had contributed to strip Gierke’s theory of its some of its “spectral attributes,” thereby “offering a working basis for the solution of corporate problems.” As “the true nature of the corporation [was] gradually becoming plainer” (Davis, 1905: 267), the “dogmatic speculations” (Gray, 1909: 54) associated with

²⁴² Freund’s theory received some attention mainly after Maitland’s translation, although his innovative work was not widely cited (see Harris, 2006: 1435; Gindis, 2007: 275). In an early endorsement, Pepper’s (1897: 479) review of Clark’s (1897) treatise based on the aggregate view of the corporation invited readers to study the more satisfying work by Freund.

²⁴³ Robinson’s observation is in line with Freund’s (1904: 356) overall position that the status of unincorporated associations is in many key respects “uncertain and unsatisfactory” because “the right to hold property in a corporate capacity” cannot be guaranteed by “contractual stipulations between the members.”
Gierke came to be used as a foil by corporate theorists concerned mainly with practical questions about corporate rights and duties. For many, this meant rejecting both concession and aggregate theories while acknowledging the key role of the policy and, by extension, of legislation (see Gray, 1909: 54-55; Machen, 1911: 361).244

These considerations played an important role in the vivid debates surrounding the legal status of unincorporated associations such as trade unions and partnerships.245 While the judicial recognition of the mercantile or “entity” view of partnerships as de facto legal persons was widespread (see Burdick, 1902: 217ff; Cowles, 1903: 343ff; Brennan, 1904: 208), it was not uniform, and this created some uncertainty, particularly in matters of interstate commerce. Responding to this situation, the drafters of the Uniform Partnership Act of 1914 attempted to agree on “such fundamental matters as the legal nature of a partnership, the rights of the members in partnership property, or even their relation to third persons” (Lewis, 1911: 100), but failed to do so, according to some observers, in a logically coherent manner (Crane, 1915: 789; Rowley, 1916: 128ff). Indeed, as one outspoken critic

244 See Freund (1906) on the important links between jurisprudence and legislation.

245 The issue of the legal status of trade unions was particularly debated in the aftermath of Taft Vale (see Brandeis, 1903; Wambaugh, 1903; Anon, 1903). Brandeis (1903: 13) famously argued that unions should incorporate, becoming legally responsible for the actions of their members, as this would enhance the unions’ legitimacy in the eyes of employers, thereby reducing the “bitter antagonism” between capital and labor. Like their British counterparts, however, Gompers and other American labor leaders rejected this idea (see Meltzer, 1998). See Adelstein (1989) for a detailed discussion of Brandeis’s use of corporate theory before and after his appointment to the Supreme Court in 1916.
pointed out, by refusing to formally define the partnership as a legal person, the drafters not only fell short of their stated objective, namely to link varied business practices to a single statutory form, but also failed to achieve consistency with existing legislation, most notably the Bankruptcy Act of 1898 (Crane, 1915: 769; 1916: 842). 246

Overall, to borrow Justice Holmes’s famous expression in the 1908 case of Donnell, the new partnership statute interposed only a partial “nonconductor” between the partnership property and the partners: as Holmes himself put it a few years later in Francis, “partnership debts are debts of the members of the firm.” The separation of control and interest emphasized by Freund is accordingly incomplete. Contrary to partnerships, the corporation’s legal personality completely insulates corporate property from shareholders and their creditors, 247 and this is the legal instrument that uniquely fosters the corporation’s unity and continuity over time. The resemblance between

246 The Bankruptcy Act clearly construed partnerships as falling under the category of “persons” alongside corporations (see Anon, 1908: 391; Hough, 1908: 603; Burdick, 1909: 396). The text of the Uniform Partnership Act was highly ambiguous in this respect since it recognized that partnerships as such could hold property but at the same time that partners were co-owners of this property. As Drake (1917: 626) put it, the statute recognized “the composite entity of the group and not the unit entity of an extrinsic juristic person.” This misalignment between formal statute and business practices was corrected only in the Revised Uniform Partnership Act of 1997, that defined the partnership as “an entity distinct from its partners” (§201a).

247 Elsewhere, Holmes argued that “the corporation is a person, and its ownership is a non-conductor that makes it impossible to attribute an interest in its property to its members” (quoted in Berger, 1955: 809).
Freund’s account and the notion of “asset partitioning” suggested by Hansmann and Kraakman (2000a: 393ff; 2000b: 810ff) is striking. Hansmann, Kraakman and Squire’s (2006: 1337ff) discussion of “entity shielding,” whereby partnerships benefit from “weak entity shielding” and corporations from “strong entity shielding,” can be traced back over a century earlier to Freund’s emphasis on what Blair (2004) called the “neglected benefits” of entity status.
Chapter 6. Hamlet with the Prince

My contribution to economics has been to urge the inclusion in our analysis of features of the economic system so obvious that … they have tended to be overlooked.

Coase (1992: 713)

Freund was truly a precursor of the modern law and economics of the firm. Unfortunately, Freund’s insights were lost in the subsequent debates leading up to Berle and Means’s (1932) classic *The Modern Corporation and Private Property*. Many of Freund’s contemporaries, including luminaries such as Hohfeld, continued to view of the notion of corporate personality with great suspicion, and argued that courts should ruthlessly pierce the corporate veil whenever necessary. As Hohfeld (1909: 289) put it in his influential essay on shareholder liabilities, “when all is said and done, a corporation is just an association of natural persons.” In many circles, commentators derided the Continental debates about the nature of corporate personality (Seymour, 1903: 529ff; Singleton, 1912: 291f; Wormser, 1912: 496), and with the rise
of the legal realist movement that favored empirical rather than conceptual jurisprudence (Douglas, 1929; Radin, 1932; Cohen, 1935), the debate progressively died out.\(^\text{248}\) As jurists elected to leave the “heaven for legal concepts,” to borrow Jhering’s (1985 [1884]) mocking phrase, by the late 1920s it was effectively dead. Not, as Derham (1958: 1) observed, “because the argument was won or lost or the problem resolved, but because the sweep of events moved on.”

A legally-grounded view of the firm, linking the theory of the firm literature and the insights revealed in the preceding autopsy of the corporate personality controversy, is proposed in what follows. The issue of legal personality has been the source of so much confusion that it is crucial to demonstrate that its inclusion in the theory of the firm does not amount to reification. Much of the controversy about the nature of legal personality, both past and present, is due to some recurrent misunderstandings and conflations that can be avoided once the precise meaning of personhood is understood. Although there are now some works about the importance of legal personality (Spulber, 2009a; North, Wallis and Weingast, 2012; Van den Steen, 2010), the reception of the notion among economists is all but non-existent. Hart’s (2011) review of Spulber reveals the lasting and damaging effect of Jensen and Meckling’s rejection of personification. Chapter 6 begins by distinguishing three notions of personhood relevant to any discussion of legal personality such that the reification illusion is avoided (6.1). The resulting conception is shown to be relevant for our

\(^{248}\) Detailed discussions of the legal realist movement can be found in Duxbury (1995) and Schlegel (1995).
understanding of the institutional structure of production (6.2), and a starting point for future research (6.3).

6.1 The meaning of personhood

Historically, the controversy surrounding the nature and origins of legal personality has been implicitly or explicitly linked to the issue of the nature and origins of legal rights, and more specifically to the question of the locus of legal rights and duties. These questions revolve around the fundamental distinction between “subjects” and “objects,” or “persons” and “things,” a distinction that goes back at least to Roman law, and is an indispensable building block of every system of law and political economy (see Allen, 1940: 437; Cairns, 1941: 96; Pound, 1959: 266; Radin, 1982: 957; Ellerman, 1988; Iwai, 1999: 587; Davies and Naffine, 2001: 24; Trahan, 2008: 9). As Blackstone (1766: 16, emphasis in original) wrote, the fact that “the objects of dominion or property are things, as contradistinguished from persons,” cannot be disputed, and the classical aphorism attributed to Gaius, that law pertains to persons, things, and actions, remains true today. Arguably, since persons and things are basic categories of what Smith (2004: 8ff) has called

\[ \text{249 The distinction was present in the Code of Hammurabi, one of the oldest written codes of law in recorded history (see VerSteeg, 2000: 65ff).} \]
law’s “ontological inventories,” decisions concerning who or what counts as a “person” from the legal point of view have profound implications.250

The term “person” derives from the Latin persona meaning “actor’s mask.”251 This is why the Roman lawyer found it suitable to denote the subject of civil rights and duties. As Hallis (1930: xix) explained, a person “was one who could be a party in a legal dispute, one who could, so to speak, act in a legal drama” (see also Thorburn, 1917: 308; Duff, 1938: 3ff; Fuller, 1967: 19; Derham, 1958: 12ff; Stoljar, 1973: 5; Hollis, 1985: 219; Teichman, 1985: 177; Poole, 1996: 39).252 Accordingly, legal persons were the actors that courts gave legal standing to as dramatis personae. Not all human beings, however, had legal standing. In early Roman civil law, for instance, “a slave was a res” (Buckland, 1908: 3; see also Watson, 1967: 173), a thing that was owned, rather than a persona, and therefore did not have the right to own property or initiate any legal actions. Although in late Roman law slaves acquired some limited rights (see Johnston, 1999: 43), it was clear that they were not persons in the then legally meaningful sense of possessing freedom,

250 The legal mind, as Vining (1978: 143) observed, will always look to “the persons that populate the legal world.” It follows, to use MacCormick’s (2007: 77) words, that it is difficult to gain “an understanding of law and how it works” without some “reflection on the idea of a person.”

251 According to Dejnožka (2007: 45-46), the term “person” could refer to the mask worn by an actor, the actor behind the mask, or the actor-in-the-mask.

252 Closely related to the notion of person was the the concept of capacitas that referred to “a status conferred upon citizens for the purpose of enabling them to participate in the economic life of the policy” (Deakin, 2006: 318).
citizenship or family rights (see Stein, 1999: 19; 2007: 22; Mousourakis, 2007: 118). This conception of personhood as legal identity based on social status is not confined to the distant past but is central to any discussion of legal personality (see Pound, 1959: 262ff).

The original accent on the social face that each member of society wears in a legal or public forum, that List and Pettit (2011: 171) call the “performative conception of personhood,” contrasts with the definition of the person as an “individual substance of a rational nature” given in the sixth century by the theologian Boethius (cited by Marshall, 1950: 472; see also Hallis, 1930: xx; Teichman, 1985: 175). In fact, as Groarke (2010: 299) observed, “the most striking philosophical feature of the Boethian person is that the private and the public components of the person are separated from each other.” A comparable position can be found in Locke’s (1975 [1690]: 335-336) claim that “person stands for … a thinking, intelligent being,” and that “consciousness makes personal identity” not only at any point in time.

---

253 Significantly, over the course of the nineteenth century, the abolition of slavery (1833 in the British Empire and 1865 in the United States) and the abolition of coverture for married women (1880s in the United Kingdom and the United States) both involved the acquisition of separate legal personality. Before their emancipation, these categories of human beings could not own property or be parties to contracts in their own right. In addition, the legal personality of various groups of foreigners also had to be progressively defined in each jurisdiction (see Mark, 2001: 8647).

254 This definition was endorsed by Aquinas in the thirteenth century, giving it “almost authoritative standing” (List and Pettit, 2011: 171), precisely at the time of Innocent IV’s famous definition of the corporation as a persona ficta. In other words, Innocent was denying that the corporation could be a Boethian person.
but also over time (see Davis, 2003: 5-6). Although Locke also argued that “person … is a forensic term appropriating actions and their merit” that “belongs only to intelligent agents capable of a law” (Locke, 1975 [1690]: 346), he was less concerned with the legal practices of ascribing actions and their merit to specific individuals than with the conditions under which these practices would make sense at all (see Poole, 1996: 40).

Distinguishing the “metaphysical notion” of the person as a conscious, rational agent from the “moral notion” of the person as an accountable agent possessing rights and responsibilities, Dennett (1976: 176ff) proceeded to examine the conditions under which the two notions overlap. Clearly, he pointed out, there are “conditions that exempt human beings from personhood, or at least some very important elements of personhood” (Dennett, 1976: 175). Indeed, the rights and duties of several categories of human beings that are not in full possession of their mental capacities are drastically diminished. Hence, infants, minors, the mentally disabled and those declared insane normally cannot be parties to contracts, make legally binding decisions by themselves or be held fully accountable for their actions (see Honoré, 1999: 17-18; Moore, 2000: 843; Cane, 2002: 143ff). Of course, this is not to deny that even when their actions cannot be imputed to them for these valid reasons, all human beings are “persons” not only in common

---

255 Like the Boethian person, Locke’s person is an autonomous and private being, confined within a “first-person world” (Davis, 2003: 6).

256 This led Dennett (1976: 190) to conclude that “the concept of a person is … inescapably normative or idealized.”
parlance but also in the broad Kantian sense of “ends in themselves, i.e., as something that may not be used merely as means” (Kant, 2002 [1785]: 46).²⁵⁷

These considerations reveal that there are three notions of personhood that need to be carefully considered in any discussion of legal personality.²⁵⁸ First, there is the ordinary language view that “person” and “human being” are interchangeable or co-extensive (see Teichman, 1985: 177). This belief is consistent with the broad Kantian sense, and is axiomatic in discussions of “human rights” in which the fact that human beings are the “natural” subjects of rights from birth onwards by the mere fact of being born human, irrespective of considerations regarding their mental or physical state, is taken for granted (see Naffine, 2003: 361; Ohlin, 2005: 234; Chwaszcza, 2010: 333).²⁵⁹ For proponents of this approach, the quality of personhood is not attributed by law to human beings (see Beitz, 2009: 52-53). On the

²⁵⁷ As opposed to the thing, for Kant the person is a “rational being” (Kant, 2002 [1785]: 46), namely “a subject who is capable of having his actions imputed to him” (Kant, 1887 [1796]: 31-32). Elsewhere, Kant pointed out that “a person cannot be property and so cannot be a thing that can be owned, for it is impossible to be a person and a thing, a proprietor and the property” (cited in Davies and Naffine, 2001: 2).

²⁵⁸ A useful discussion of these three types is provided by Naffine (2003: 349ff; 2004: 626ff). See also Davies and Naffine (2001: 57ff). In the same vein, Barker (1934: lxiii-lxiv) distinguished three notions of personhood.

²⁵⁹ More precisely, Article 6 of the Universal Declaration of Human Rights adopted by the United Nations in 1948, stating that “everyone has the right to recognition everywhere as a person before the law,” is intended to ensure that every human being is treated as a subject rather than an object of law in every jurisdiction (see Jayawickrama, 2003: 595-596).
contrary, as natural law theorists have long argued, positive law is subordinated, to quote Del Vecchio (1920: 132-133), to the law “which has its foundation in human nature” (see also Finnis, 1980: 23ff; George, 1992: 32ff). As Finnis (2000: 11ff) has emphasized, human beings always have ontological, explanatory and moral “priority.”

As if to illustrate the long-standing dispute between natural law and legal positivism, most if not all jurisdictions clearly distinguish between the human being and the “responsible subject” (Naffine, 2004: 628). In fact, many actual legal rules concerning property, contract, standing, accountability, and so on, rely on this second view of the person, even though this more restrictive definition allows only a subset of human beings – those of a certain age and in sufficient possession of their faculties – to be fully admitted into the legal realm.260 A person thus defined is truly a “moral agent [that] can be both morally and legally accountable for his action because his actions are guided by reason: he knows what he is doing and still chooses to act as he does” (Naffine, 2003: 362). The intuitively appealing link between moral responsibility, free will and agency, that Frankfurt (1969: 829) called “the principle of alternate possibilities,” has led many to see in the responsible subject the “ideal legal actor” (Naffine, 2009: 67) or the “default legal person” (Blumenthal, 2007: 1138).

260 Although all humans may have some human rights (such as the right to life) from birth onwards, a human being is not born with civil rights and legal capacities (see McHugh, 1992: 458). Legal personality is acquired by degree: the typical adult becomes a full-blown legal person “by acquiring rights, powers, and duties, which gather cumulatively” (Tur, 1987: 123) over time.
Much of the confusion surrounding the issue of the firm’s legal personality is due to the tendency to address the matter with only these two, all too often conflated, definitions of personhood in mind. The inclination to frame the debate exclusively in terms of what Rovane (1998: 136) appositely calls “human-size persons” has obscured the fact that there is a third and more useful possibility. When the term “person” is defined in line with its original meaning as “mask” worn in the legal drama, it is easy to see that “it is [only] the legally-endowed capacity to attract legal relations, and hence bear rights and duties, which defines the person” (Naffine, 2003: 366; see also Naffine, 2009: 30).\footnote{From this perspective, Kelsen (1945: 99) defined the legal person as a “point of imputation” of rights and duties that arise in legal relations (see also Kelsen, 1992 [1934]: 50), while Kocourek (1927: 57) claimed that “legal persons … in a scientific sense are conceptual points of reference” in the analysis of legal relations, and by which law adjusts human interests (see also Kocourek, 1922: 517; Warren, 1929: 841ff). This was essentially Freund’s (1897: 81) view.}

Accordingly, this definition severs the misleading link between the flesh-and-blood human being and the legal person (Naffine, 2003: 350; 2004: 626).\footnote{The link between personhood and the will theory of rights is also usefully severed.} Properly understood, as Tur (1987: 121) explained, the concept of legal personality is “wholly formal” (see also Kocourek, 1927: 292; Heilman, 1927: 204; Smith, 1928: 293-294; Rutledge, 1929: 351; Nékám, 1938: 26; Lawson, 1957: 915; Derham, 1958: 5; Pound, 1959: 261), with the implication that legal persons may be “of as many kinds as the law pleases” (Salmond, 1902: 344; see also Pollock, 1911b: 110-111).
This is the conception of the legal person that opens the gate to the legal realm to candidates that would otherwise be excluded. On this view, there is no essential difference between the legal personality attributed under certain conditions to human beings and the legal personality attributed under different conditions to groups of human beings or organizations such as firms. In both cases, legal personality is not “natural” if this term is taken to mean “pre-legal,” as some of Maitland’s followers had pointed out (see Salmond, 1902; Brown, 1905; Carr, 1905; Geldart, 1911). Like the legal realm itself, legal personality is always “artificial,” to paraphrase Machen (1911: 17), but this does not mean that it is “fictitious” (see also Dewey, 1926: 655, n1; Smith, 1928: 293; Fuller, 1967: 13; Dejnožka, 2007: 7). Adopting this view does not imply that the attribution of legal personality to any inanimate thing, animal or association of human beings is wholly

---

263 Alongside firms of various types, this includes idols, temples, ships, foundations, municipalities, states, regional organizations, and international organizations (see Duff, 1927; Smith, 1928; Wolff, 1938; Aufricht, 1943; McDougal, Burke and Vlasic, 1960; Frug, 1980; Collier, 1992; Tiunov, 1993; Paasivirta, 1997; Panico, 2012). Further extensions to trees, animals, species, artificial intelligences, computers, and electronic agents have been proposed (see Stone, 1972; Varner, 1987; Solum, 1992; Allen and Waddison, 1996; Teubner, 2006).

264 Dewey (1926: 657) and Barker (1934: lxxvi) also made this point. In Deakin’s (2012a: 115-116) words, “it is no more a ‘fiction’ to assign legal personality to organisational structures than it is to grant it to natural persons.”

265 As Koessler (1949: 449) put it, “speculations about the reality or unreality of corporate personality ... have no more sense than speculations about the reality or unreality of the conception of property or of other established institutions of a legal nature.”
unproblematic. But it does create the conditions in which a discussion of these issues is possible without the undesirable emotional associations that often plague the debate.

The history of corporate theory is replete with examples of emotionally-charged and often politically-motivated uses of the notion of personhood. In fact, a recurring aspect of past and present disputes is that both proponents and opponents of corporate regulation have ably exploited the ambiguity of the term “person,” and have used the strategy of equivocation to justify their normative agendas. As a result, to quote Dewey (1926: 670), “the doctrine that true personality resides only in the ‘natural’ person has been worked in opposed directions.” Indeed, the view that “corporations are not persons” lies at the foundations of Nader’s criticism of the Supreme Court’s protection of big business and his related call for standards of corporate behavior (see Nader, Green and Seligman, 1976a: 183; 1976b: 207; Nader and Meyer, 1988: 31; Nader and Weissman, 2010: 19). At the same time, however, it is precisely the argument that the “personalization of the firm … is seriously misleading,” as Jensen and Meckling (1976: 311) put it, that is mobilized to

---

266 An interesting feature of the currently unresolved debate surrounding animal rights is that the protagonists hold contrasting views of legal personality. Whereas Wise (2004: 25ff) and others argue that at least some highly evolved animals should be formally recognized as legal persons, precisely because they are (or seem to come close to) sentient beings, opponents claim that the only legitimate legal persons are human beings (e.g., Epstein, 2004: 151).

267 In a related deliberately ambiguous argument Jensen and Meckling (1976: 306, n.6; 1978: 36; see also Meckling and Jensen, 1977: 41) claimed that all rights, particularly property rights, are “human rights.”
dismiss Nader’s proposals by members of the anti-regulatory law and economics movement (e.g., Hessen, 1976a: 14; 1979b: 40ff; Fischel, 1982a: 1273; Klein, 1982: 1523).  

Strikingly, both sides of this debate claim that firms are legal fictions, yet both sides appear to accuse the other of upholding the exact opposite, that is, of falling prey to the reification illusion. Of course, neither side comes anywhere near to accepting the Gierkean idea of the corporate “real person,” defined as a “psycho-spiritual living unity” with a will of its own (Gierke, 1935 [1902]: 152). There is something suspicious about Gierke’s theory because it transforms what could in some respects be deemed as an interesting analogy between associations and human beings into something more. It is one thing to claim that when the bond of association is strong a group of human beings can be properly described as a singular decision-making unit. It is quite another to argue that the group is physically a living

---

268 As Flynn (1987: 137) explained, concepts such as corporate personality can be “manipulated as devices to hide and give effect to underlying ideological preferences.” Samuels (1987: 115ff) similarly argued that a case can be made for the fact alternative conceptions of corporate personality express the changing belief systems that accompany the transformation of the power structure of the economy. See Millon (2001: 51ff) for a more detailed discussion of what Smith (2001: 69) has appropriately called the “opportunistic use of the idea of corporate personality.” A conspicuous illustration was recently provided in the United States following Citizens United, where the unprecedented yet misguided grassroots campaign to “abolish corporate personhood” (e.g., [https://movetoamend.org](https://movetoamend.org)) intensified when Romney, implicitly referring to the view associated with Jensen and Meckling, commented in August 2011 that “corporations are people too” (see [http://www.nytimes.com/2011/08/12/us/politics/12romney.html](http://www.nytimes.com/2011/08/12/us/politics/12romney.html)).
unit (see Barker, 1934: xxix). From this point of view, Gierke clearly committed what Ryle (1949: 16ff) famously called a “category mistake.” Although Maitland was generally more careful in his formulations, his claim that “if n men unite themselves in an organised body, jurisprudence, unless it wishes to pulverize the group, must see $n + 1$ persons” (Maitland, 1905: 198), did little to resolve the ambiguities of the corporate personality controversy.

These and other misunderstandings can be avoided if the notion of personhood is disassociated from both the flesh-and-blood human being and the will theory of rights, and the original sense of the term is adopted. When the law treats the firm as a “person” nothing more than the fact that the firm has a point of imputation for rights and duties that arise in legal relations should be implied. It is important to understand, as Dewey (1926: 656) explained, that

what “person” signifies in popular speech, or in psychology, or in philosophy or morals, [is] as irrelevant, to employ an exaggerated simile, as it would be to argue that because a wine is called “dry,” it has the properties of dry solids; or that, because it does not have those properties, wine cannot possibly be “dry.”

---

269 The attempt to literally endow the group with life was, as Teubner (1988: 134) put it, “Gierke’s cardinal error.” Indeed, “to live or not to live … cannot be the question, the keystone to the nature of juristic personality” (Wolff, 1938: 504).

270 A category mistake can involve presenting a thing as being of certain kind when in actual fact it belongs to another, or attributing a property to something that the thing it is attributed to logically cannot have.
Obviously, “dry” as applied to a particular wine has the kind of meaning, and only the kind of meaning, which it has when applied to the class of beverages in general. Why should not the same sort of thing hold of the use of “person” in law?

Perhaps, as Nékám (1938: 67) suggested, the typical use of the idea of personality to refer to a legally-endowed subject of rights and duties is inappropriate, and “the term legal entity is much better” (Nékám, 1938: 70). Or perhaps, as Pound (1959: 261) similarly argued, the expression “legal unit” is preferable since legal persons are ultimately the “units of the legal order.” Whatever the label, the key point to acknowledge is that only the more formal view of the nature of legal persons is broad enough to accommodate the large variety of cases to be found in the real world. It follows that the adoption of this meaning of personhood, that helps “describe with simplicity and accuracy all the relevant phenomena of the legal system” (Nékám, 1938: 70), is inevitable. This broad definition is compatible with the fact that who or what is regarded as a legal person, and the number and quality of the rights and duties thus recognized, depend on the “changing evaluation of the given community” (Nékám, 1938: 116). This historical contingency equally characterizes human beings and firms.

---


272 In the same vein, Fuller (1967: 14) observed that the term “legal subject” may be preferable.
6.2 The firm, the market and the law

This thesis has argued that in light of the basic principle to be found in any jurisdiction that only legal persons may own property and have the capacity to contract, and the implication that legally enforceable contracts can only exist between legal persons, it is something of a paradox that legal personality is absent from the dominant narrative in the contractual theory of the firm. Indeed, many of the things that theorists of the firm purport to explain simply cannot be accounted for when legal persons missing: without the device of legal personality, firms would not be able to own and pledge assets, raise finance, contract with one another, merge, and act in other market-like ways, not to mention access the legal system in the case of a dispute. It follows that a theory of the firm that overlooks these elementary institutional facts about the firm is like Hamlet without the Prince. Not surprisingly, it was argued, Holmström (1999: 75) found himself unable to account for “one of the most significant and robust empirical regularities to be explained by any theory of the firm,” namely that “economic contracts are made with firms, not their employees or owners.”

As we have seen, Holmström’s puzzle stems from the erroneous assumption that the efficiency-enhancing central contractual agent, identified by Coase as a core feature of the firm, is a flesh-and-blood human being or a group of such human beings, rather than the separate legal entity in which ownership rights over assets used in production are vested, and in whose name contracts are made. This mistaken assumption was inherited from Coase (1937) himself, and characterizes later stylized accounts of the emergence and development of the firm in a market economy (e.g., Alchian
On this view, the most basic firm involves an entrepreneur, who owns or controls access to the assets used in production, holds residual claim rights to the net value of the joint output, and is the only common signatory of a set of contracts with employees, suppliers, customers, and other agents. The subsequent separation of ownership and control merely means that managers, who are also employees, control access to the assets and play the role of the central contractual agent, acting in the entrepreneur’s name. This setup makes it difficult to come to terms with the significant and robust empirical regularity highlighted by Holmström.

Unaware of this problem, theorists of the firm have concentrated on Coase’s second claim, but a consensus regarding the boundaries the firm was never reached. In fact, the organizational anonymity problem, or the apparent irrelevance of the institutional setting for contracting originally identified by Holmström and Tirole (1989: 69) in Alchian and Demsetz’s (1972) story, was never solved in a satisfactory manner, even after the notion of asset specificity had fundamentally transformed the theory of the firm narrative. As Holmström (1999: 87) argued, the same problem cropped up in the Hart and Moore (1990) view of the firm, and it is equally clear that the issue was far from resolved in Rajan and Zingales (1998) account of the human-capital intensive firm. In all of these cases, there is no obvious way to determine whether a given contractual relation is inside or outside the firm, or whether it involves one or two firms, as Cheung (1983: 17) put it. From this point of view, Williamson’s (1991a) discussion of discrete structural alternatives, that highlights the different dispute-resolution mechanisms available to contracting parties in different institutional settings, is the only theory of the
firm that tackles the organizational anonymity problem head-on. Significantly, Williamson appealed to legal concepts and practices.

However, Williamson stopped short of including legal personality in his account, even though it is precisely because firms have standing in courts as singular parties (see Iacobucci and Triantis, 2007: 524) that the notion of forbearance contributes to an understanding of firm boundaries. Since all forms of market and hybrid governance structures always involve at least two distinct legal persons, contrary to the unified governance structure of the firm, legal personality can more generally help identify the distribution of transactions. As we have seen, the benefits of including the firm’s legal entity status in the theory of the firm narrative are evident, and the implicit recognition of this fact can be found scattered in the literature. Importantly, as Demsetz (1997: 429) pointed out, a case can be made for the fact that the continuity and integrity of the firm are enhanced when the firm itself owns the assets used in production. This fits Rajan and Zingales’s (1998) argument about the efficiency-enhancing properties of third party ownership if, as Blair (1999: 85) has argued, the third party owner is viewed as none other than the separate entity created for the firm by the legal system.

Since legal personality is clearly a transaction cost-reducing device (see Spulber, 2009a: 63ff) that is one of those “features of the economic system so obvious that … they have tended to be overlooked,” to paraphrase Coase (1992: 713), the thesis set out to investigate the reasons behind this state of affairs. Alongside the widespread view among economists that firms can be defined with little or no reference to law (e.g., Cheung, 1983; Alchian, 1984; Demsetz, 1988; Williamson, 1990a; Rajan and Zingales, 2001), based on the equally pervasive belief that “legal concepts … are ‘fictions’ which are liable
to conceal the true nature of the [economic] forces at work” (Deakin, 2012b: 344), the lasting influence of Jensen and Meckling’s (1976) ambiguous dismissal of legal personality was identified as the principal impediment to its inclusion in the theory of the firm narrative. Indeed, their portrayal of the “personalization” of the firm as a misleading legal fiction that fuels nonsensical talk of the firm’s objectives and responsibilities (Jensen and Meckling, 1976: 311) has had a damaging effect.

In order to reveal just how misguided and overblown this warning about the reification illusion actually was, Jensen and Meckling’s discussion was contextualised in the 1970s debate on how to tame the giant corporation (see Nader, Green and Seligman, 1976). Nader and other critics of big business claimed that corporate features such as legal personality were state-created privileges, and that this justified the state’s regulation of corporate rights and duties by imposing standards of socially responsible corporate behavior. Jensen and Meckling targeted this view when they emphasized the contractual essence of the firm, the specific features of which had “survived the market test against potential alternatives” (Jensen and Meckling, 1976: 357), and the fact that all rights were “human rights” (Jensen and Meckling, 1976: 307, n.6; 1978: 32). Their arguments provided a theory of the firm foundation and a rallying rhetoric for members of the anti-regulatory corporate law and economics movement developed in Manne’s footsteps (e.g., Hessen, 1979a; Posner and Scott, 1980; Fischel, 1982a; Klein, 1982; Baysinger and Butler 1985a), who picked up and amplified the explicit link between legal personality, the reification illusion and inefficient state involvement in private business.
Accordingly, if legal personality is to be rehabilitated, this link, that reflects a normative agenda rather than undeniable institutional facts, needs to be severed. In view of showing that the notion of legal personality need not imply anthropomorphism and state regulation of business, the debate was put into historical perspective. The thesis examined the corporate personality controversy of the past, demonstrating that although corporate personality was indeed widely perceived to be a concession of state powers throughout most of its tortuous history, particularly before the liberalization of incorporation laws in the late nineteenth century, very few corporate theorists fell prey to the anthropomorphic fallacies that Jensen and Meckling presented as inevitable. If one excludes Gierke’s (1900) talk of “real corporate personality,” and some of Maitland’s (1905) more ambiguous statements, most corporate theorists discussed in this thesis generally avoided category mistakes of this kind, even when they advocated some kind of state regulation in response to abuses of corporate power that accompanied the rise of big business at the turn of the twentieth century.

The clear view that emerged in Britain and America, as discussions focused not only on the nature of corporations but also on the practical implications of the extension of legal personality to previously unincorporated associations such as partnerships and trade unions, was that corporations were neither state concessions nor merely private contracts (e.g., Davis, 1897; Freund, 1897; Williams, 1899; Pepper, 1901; Brown, 1905; Geldart, 1911; Machen, 1911; Laski, 1916a). It became increasingly acknowledged that dubious anthropomorphic statements stemmed from the traditional will theory of rights, and could readily be avoided by appealing to an interest theory of rights (see Salmond, 1902; Gray, 1909). As soon as rights were separated from volition, it became clear that the notion of
corporate rights and duties was no more “abnormal and illogical,” as Freund (1897: 48) put it in what could properly be considered as the first rational study of the value of corporate law, than individual rights and duties. In both cases, as Freund (1897: 81) argued, the term “person” denotes nothing more than a distinctive party to legal relations.

When personhood is defined in line with its original meaning, that is, when the other senses of the term, those derived from ordinary language or psychology or morals, are set aside as irrelevant following Dewey’s (1926: 656) recommendation, it is easy to see that a discussion of the value of legal personality, without any emotionally-charged and politically-motivated equivocations, becomes possible. From this point of view, contrary to some of his contemporaries who attempted to avoid the notion of corporate personality for the same misguided reasons as Jensen and Meckling (e.g., Morawetz, 1882; Taylor, 1884; Trapnell, 1897; Clark and Marshall, 1901; Purdy, 1905), Freund identified the separation of control and interest uniquely achieved by the assignment of legal personality as the principal service rendered by the legal system to an association of resources owners seeking to make a joint surplus. By creating a “property-holding body” (Freund, 1897: 9) entirely distinct from all the human beings involved, and by attributing undivided control over this property to a governing body whose position cannot be revoked, the law provides “security both against outsiders and against defection on the part of the members” (Freund, 1897: 22) that cannot be otherwise reliably achieved. Different legal forms support different degrees of security, but separation of control and interest is common to all these forms.
By emphasizing the neglected benefits of entity status, to borrow Blair’s (2004) expression, Freund demonstrated that substance could not be easily divorced from form, anticipating some of the recent work in this area (e.g., Blair and Stout, 1999; Hansmann and Kraakman, 2000a). Taken together, these ideas can be used to rehabilitate the notion of legal personality, and reverse the tendency to downplay or ignore the role of law in constituting and defining the firm (see Behrens, 1985; Franke, 1987; Masten, 1988; Hodgson, 2002). Arguably, the “legal-economic nexus” discussed by Samuels (1989), namely the fundamental “interrelations between legal and economic processes” (Samuels, 2007: 4), cannot be ignored. Given that it is difficult to understand the institutional structure of production without acknowledging, as Coase (1992: 717-718) put it, that “the legal system will have a profound effect on the working of the economic system and may in certain respects be said to control it” (see also Coase, 1988a: 10), it is clear that a legally-grounded view of the firm must take into account “what is known empirically of the way legal systems constitute and regulate the business enterprise” (Deakin, 2012b: 347-348).

From this perspective, whether it is set up by a single entrepreneur or by an association of resource owners, it is clear that the firm comes into existence following a registration or incorporation procedure, and becomes fully operational with the acquisition of legal personality, namely the legally endowed capacity for property, contract and litigation that allows what is fundamentally a specialized economic undertaking “to rank as a unit in the legal scheme,” as Carr (1905: 185) put it. The constitutive procedure that creates a singular point of imputation for rights and duties, entirely distinct from all of the human beings involved, that acts as a “nexus for contracts” (Hansmann, 2013: 892, emphasis in original) literally carrying the
organizational framework of the firm (see Schanze, 2006; Deakin, 2012a), is truly what Spulber (2009a: 152) described as a “foundational shift.” The value of operating as a legal entity lies in the separation that it introduces between the assets locked-in to guarantee the firm’s contractual commitments, and the human beings involved. This allows the firm’s commitments to survive changes in the firm’s membership. Armed with this device, the firm’s founders can pledge assets, raise finance and do business in the firm’s own name, such that some even thrive.

A legally-grounded view of the firm must recognize, as Deakin (2009: 41) observed, that “many of the rules relating to the business enterprise … are statutory in origin.” This fact points to the limitations of theories of the firm based entirely on the “lens of contract” and the corresponding “private ordering” framework (see Williamson, 2002a; 2002b). That this framework has generated numerous insights is undeniable, but this has come at the cost of underplaying the role of statutory law and the state more generally in the emergence and rise of the firm in history. This is remarkably clear when it comes to the issue of legal personality. Market forces and network externalities play an important role in the diffusion and adoption dynamics of the institutional arrangements defining the firm, but this does not mean that the legislator’s role can be explained away or otherwise ignored. Indeed, considerations of “public ordering” are needed from both the analytical and historical perspectives to account for the “emergence of new organizational species” (Pagano, 2011: 379; see also Pagano, 2010: 121ff; 2012: 1272) such as firms. Hence the idea that “in the beginning there were markets” (Williamson, 1975: 20) should not be accorded more weight than it deserves.
Although it is true that some elements of entity shielding, such as an order of priority in bankruptcy, had been achieved by some unincorporated associations before the corresponding legislation and the common availability of the corporate form (see Lamoreaux, 1995; Getzler and Macnair, 2005), it is equally true that legislation enacted over the nineteenth and early twentieth centuries standardized varied business practices, significantly reducing the transaction costs involved in undertaking business operations at a larger scale and over greater distances. These changes in the institutional environment created favorable conditions for the accumulation of the specific assets that came to define advanced industrial economies (see Blair, 2003; Hansmann, Kraakman and Squire, 2006; Lamoreaux and Rosenthal, 2006). Thanks to the “institutional glue” (Gindis, 2007: 279; 2009: 40) provided by legal entity status and other complementary institutions, increasingly bigger and far-reaching things did not fall apart; their centers held with ever-increasing amounts of locked-in capital.

The firm’s expansion through vertical integration depends on “make-or-buy” decisions that contribute, as theorists of the firm have long argued, to the definition of firm boundaries. Nonetheless, the presentation of firms and markets as substitutes is misleading because it encourages the conflation of markets and market transactions. Market transactions are events and relations between firms and other economic actors, as Simon’s (1991: 27) “visitor from Mars” would concur. Markets, however, are not reducible to market transactions since they are organized systems of property rights exchange (see Hodgson, 2002, 2008a, 2008b) in which some of the most important actors are firms. The legal structure that is essential for markets relies on the institutional fact that firms are singular legal entities that can hold property and act in certain market-like ways (e.g., engage in market transactions,
compete, merge). It is clear, from this perspective, that the difference between the firm and the market is a difference in kind rather than in degree (see Gindis, 2009: 41). There is no continuum between firms and markets. Far from being substitutes, firms and markets are complementary institutions of capitalism.

A working definition of the firm can now be suggested. The firm is neither a coalition of owners nor a collection of assets but a profit-seeking specialized production unit set up for the sale of goods or services, and endowed by law with the capacity to act as a singular legal person. This definition emphasizes both the economic entity that creates the surplus and the legal medium through which it acts, recognizing that without the device of legal personality most if not all of the firm’s activities would be difficult to sustain, if possible at all. Significantly, employment relations are not essential to this definition of the firm that is compatible with a variety of rationales for the firm’s existence, including besides transaction cost explanations those based on uncertainty, cognition and knowledge (see Hodgson, 2004b). The definition is broad enough to encompass all the observed types of firm, whether they are entrepreneurial or managerial, capitalist or socialist, nonhuman asset-intensive or human capital-intensive. At the same time it is narrow enough to exclude specialized production units that are not legally recognized, including illegal organizations, but also more complex organizational forms such as conglomerates, strategic alliances, supplier networks, all of which involve multiple firms. Arguably, it is “realistic” in the sense advocated by Coase (1937: 386) in that it corresponds to what is meant by the firm in the real world.
6.3 Future research

The preceding considerations raise a number of additional research questions, only a few of which are very briefly outlined here. A first set of issues has to do with the extension or domain of applicability of the definition of the firm suggested above. Specifically, given the particular emphasis placed on registration or incorporation, that is, given the view that a specialized economic undertaking only really becomes a fully operational firm once it is legally recognized, it seems that the definition is not applicable in the so-called “informal economy,” particularly in developing countries. Inevitably, this evokes concerns about not only the correctness of the definition but also about its usefulness. Of course, similar worries can be raised about any definition, and certainly reservations of this kind have been highlighted regarding other definitions of the firm. Indeed, these issues were at the heart of Hodgson’s (1998c: 24ff) discussion of the “Coasean tangle,” namely the lack of clarity regarding the historically specificity of the firm in Coase’s work (see also Hodgson, 2001a: 258ff).

There is a sense in which the legally-grounded view of the firm offered here does not suffer from this weakness. The objection that the definition lacks applicability in settings where the rule of law is absent or dysfunctional merely specifies the definition’s relevance in space and time. Nevertheless, the interesting issue to research is the transition from informality to formality. This question is meaningful from both the historical and contemporary perspectives. In recent times, the focus of reforms in the developing world has moved “from getting prices right to getting institutions right” (Rodrik, 2008: 100), based on the belief that development policies
should help entrepreneurs “formalize” their operations. The argument is that by facilitating access to property registries (see De Soto, 2000; Trebilcock and Veel, 2008), company registries (see Arruñada, 2010, 2012), and the like, formalization provides verifiable information that helps underpin the sort of credible commitments that are needed to foster the growth and development of firms. Of course, as North, Wallis and Weingast (2009: 150) put it, certain “doorstep conditions” must be satisfied for the transformation of personal exchange into impersonal exchange to actually work.

The investigation of these conditions from both the theoretical and empirical points of view is likely to be promising, provided that it avoids some of the sweeping assumptions made in the “legal origins” literature (e.g., Glaeser and Shleifer, 2002; Beck, Demirgüç-Kunt and Levine, 2003; La Porta, López-de-Silanes and Shleifer, 2008), namely the idea that common law systems are more conducive to business and market institutions than civil law systems. As Deakin (2009: 60) has pointed out, specific legal institutions are not “predetermined by the legacy of legal origin,” and the emergence of corporate law principles is no exception to this rule (see also Guinnane, Lamoreaux, Harris and Rosenthal, 2006: 2ff; Milhaupt and Pistor, 2008: 21ff). In fact, judging by the experience of developed economies, it always involves a complex process of transplantations and adaptations. This process can be more or less successful, and will hinge on the degree of institutional complementarity with other bodies of law, including contact law, tort law, bankruptcy law, competition law, employment law, patent law, and so on.

From a conceptual point of view, the principle that institutional complementarities are important has been firmly established, and many of the consequences have been spelled out (e.g., Pagano, 1991, 1992; Pagano and
Rowthorn, 1994; Aoki, 1994, 2001; Boyer, 2005). From a more applied point of view, although numerous specific complementarities involving various legal rules related to firms and their political environments have been examined (e.g., Pagano and Rossi, 2004; Crouch et al., 2005; Deakin, 2009; Aoki, 2010), much more remains to be done. An important but under-researched topic is the co-evolution of and complementarities between corporate personality and patent law. The crucial change that coincided with the rise of big business at the turn of the twentieth century, namely the appropriation of employee inventions by their corporate employers, has been well documented from a historical point of view (see Cherensky, 1993; Merges, 1999; Lamoreaux and Sokoloff, 2007; Fisk, 2009; Coriat and Weinstein, 2012), but more specific analytical frameworks that can help capture the multiple institutional equilibria between entity shielding rules and patent ownership rules can be usefully developed. Arguably, Pagano’s (2007) discussion of the institutional complementarities between positions, rights and duties can used as a stepping stone in this direction.


Francis v. McNeal, 228 U.S. 695 (1913)


Salomon v. A. Salomon & Co., AC 22 (1897).


Sutton’s Hospital, 77 Eng. Rep. 960 (1612).

Taft Vale Railway Co. v. Amalgamated Society of Railway Servants, UKHL 1 (1901).


Coker, Francis W. (1910), Organismic Theories of the State: Nineteenth Century Interpretations of the State as Organism or as Person, New York: Columbia University Press.


Gierke, Otto (1900), Political Theories of the Middle Age, translated with an introduction by Frederick W. Maitland, Cambridge: Cambridge University Press.


Jhering, Rudolf (1852) Der Geist des römischen Rechts, Leipzig: Auflage.


Jhering, Rudolf (1913 [1877]), Law as a Means to an End, translated by Isaac Husik with introductions by Henry Lamm and William M. Geldart, Boston: Boston Book Company.


Kantorowicz, Ernst H. (1957), The King’s Two Bodies: A Study in Mediaeval Political Theology, Princeton, NJ: Princeton University Press.


Kelsen, Hans (1992 [1934]), Introduction to the Problems of Legal Theory, translated by


Klein, Peter G. (1999), “New Institutional Economics,” in Boudewijn Bouckaert and


Kocourek, Albert (1927), *Jural Relations*, Indianapolis, IN: Bobbs-Merrill Co.


245


Mayer, Carl J. (1990), “Personalizing the Impersonal: Corporations and the Bill of
Ménard, Claude (1996), “On Clusters, Hybrids and Other Strange Forms: The Case of the


Nader, Ralph, Mark J. Green and Joel Seligman (1976a), *Taming the Giant Corporation: How the Largest Corporations Control Our Lives*, New York: W.W. Norton


Robertson, Dennis H. (1923), The Control of Industry, Cambridge: Cambridge University Press.


Robinson, Joan (1933), The Economics of Imperfect Competition, London: Macmillan.


Spence, Herbert (1851), Social Statics, London: John Chapman.


Tamanaha, Brian Z. (2006), Law as a Means to an End: Threat to the Rule of Law, Cambridge: Cambridge University Press.


Tribe, Keith (1995), Strategies of Economic Order: German Economic Discourse 1750-


Wormser, I. Maurice (1914), Private Corporations, Chicago: Blackstone Institute.

