

In Search of New Foundations

LUIGI ZINGALES*

Abstract

In this paper I argue that corporate finance theory, empirical research, practical applications, and policy recommendations are deeply rooted in an underlying theory of the firm. I also argue that although the existing theories have delivered very important and useful insights, they seem to be quite ineffective in helping us cope with the new type of firms that is emerging. I outline the characteristics that a new theory of the firm should satisfy and how such a theory could change the way we do corporate finance, both theoretically and empirically.

FOR A RELATIVELY YOUNG RESEARCHER like myself, there is a very strong tendency to look at the history of corporate finance and be overwhelmed by the giants of the recent past. A field that 40 years ago was little more than a collection of cookbook recipes that reflected practitioners' common sense is today a bona fide discipline, taught not only to future practitioners but also to doctoral students, both in business schools and in economic departments—a discipline whose ideas are now influencing other areas of economics, such as industrial organization, monetary policy, and asset pricing. The quality and the impact of the contributions that were made to the field during the last 40 years, and in particular in the period from the late 1970s to the late 1980s, justify the widespread feeling that the "golden age" of corporate finance is behind us.

Two excellent recent surveys of the main areas of corporate finance reinforce this sense: the capital structure survey by Harris and Raviv (1991) and the corporate governance survey by Shleifer and Vishny (1997). Both are very lucid categorizations of the existing literature. This lucidity is the product not only of the ability of their authors but also of the ripeness of the moment. Both surveys follow a period of intense activity in the field, and in a certain sense, they close it. It is especially noteworthy that, 10 years later, the survey by Harris and Raviv (1991) would not necessitate any dramatic rewriting. Although there have certainly been important contributions afterward, they have been mostly empirical, and they have not undermined the conceptual framework underlying Harris and Raviv's analyses.

* University of Chicago, NBER, and CEPR. Paper prepared for the 2000 AFA meeting in Boston. I benefited from the comments of Franklin Allen, Darin Clay, Francesca Cornelli, Eugene Fama, Mark Gormaise, Oliver Hart, Thomas Hellmann, Raghu Rajan, Jesus Santos, Paola Sapienza, and Andrei Shleifer.

The temptation to dedicate this piece on the state of the art in corporate finance to a celebration of past achievements, thus, is very strong, but it would fall short of the task assigned me: to provide the readers of the *Journal of Finance* with an overview of what is new and exciting in the field of corporate finance. Thus, rather than taking comfort from the great successes of the past, I will focus on the challenges the future presents. In doing so, I feel like a dwarf who has the arrogance to try to see further than his giant predecessors. In my defense, I invoke Giordano Bruno's famous line that I am a dwarf, but I am standing on the shoulders of the giants of the past. Like Giordano Bruno, I will inevitably present a very personal view of what I see from those shoulders. I hope I will not be condemned to the same fate.

The essential point of the paper is extremely simple. Corporate finance is the study of the way *firms* are financed. Theory of the firm, thus, has a tremendous impact on the way we think about corporate finance, the way we do empirical research, the policy implications we derive, and the topics we choose to study. Even the most practical applications of this discipline, like the way we value firms, are unconsciously slaves to some underlying theory of the firm.

The first goal of this paper is to show why this is the case. I will link the central issues in three main areas of corporate finance (financing, governance, and valuation) to the main theories of the firm. In doing so, I will also show that, in spite of some shortcomings, the existing theories of the firm have been quite successful in providing an intellectual backbone for the principles of corporate finance we all know and teach.

The type of firm modeled by these theories, however, was the traditional business corporation, which emerged at the beginning of the twentieth century and has since prevailed. It is a very asset-intensive and highly vertically integrated firm, with a tight control over its employees—control that is concentrated at the top of the organizational pyramid. Its boundaries are clear cut and sufficiently stable that one can take them for granted while considering the impact of financing and governance choices. In such a world, once the concept of the firm is defined, most corporate finance can be developed without a continuous reference to the underlying theory of the firm.

Not any more. The nature of the firm is changing. Large conglomerates have been broken up, and their units have been spun off as stand-alone companies. Vertically integrated manufacturers have relinquished direct control of their suppliers and moved toward looser forms of collaboration. Human capital is emerging as the most crucial asset. As a result of these changes, the boundaries of the firms are in constant flux, and financing and governance choices can easily change them. The influence of theory of the firm is no longer limited to defining the object of analysis. The interaction between the nature of the firm and corporate finance issues has become so intimate that answering the fundamental questions in theory of the firm has become a precondition for any further advancement in corporate finance.

The second and most important goal of the paper, thus, is to discuss these fundamental questions and show how addressing them might change the way we will do corporate finance in the future. I identify four such ques-

tions. The first and fundamental one is how an organization devoid of unique assets succeeds in acquiring power that differs from “ordinary market contracting between any two people” (Alchian and Demsetz (1972, 777)). An answer to this question will help us to define not only the boundaries of the firm but also the role of outside equity.

The second question is how is this power maintained and enhanced, and how is it lost? An answer to this question can explain what factors underlie a firm’s ability to capture growth opportunities or its failure to do so. Understanding these factors will not only affect the way we value firms but will also create a fundamental building block toward a theory of entrepreneurship. Entrepreneurship is the process by which new firms are created. But new firms are created to exploit growth options existing firms cannot or do not want to exploit. Thus, a theory able to explain what growth options existing firms are willing and able to exploit will also identify the opportunities for entrepreneurial activity.

The third question is how this authority-based system operates in a way different from ordinary market contracting. Not only would an answer to this question enable us to understand theoretically the effects of mergers and spin-offs, but also it will bring new life to the debate on the costs and benefits of corporate diversification.

The fourth and final question that a theory of the firm should address is how the surplus generated by the firm is allocated among its members. Such an understanding will help us derive a new approach to valuing firms that is consistent with the new nature of firms.

To illustrate concretely the benefits that can be reaped by answering these questions, I present a recent attempt in this direction and the implications that can stem from it.

The rest of the paper proceeds as follows. Section I analyzes what corporate finance is about and why it assumed this name. Section II explains why the name carried important consequences for the type of research that has been done and the type of problems that has been ignored. Section III discusses how corporate finance is deeply rooted in an underlying theory of the firm and illustrates the implications for corporate finance of the most important theories of the firm. Section IV argues that although the existing theories have delivered very important and useful insights, they seem to be quite ineffective in helping us cope with the new type of firms that is emerging. Section V outlines the characteristics that a new theory of the firm should satisfy and suggests how such a theory could change the way we do corporate finance, both theoretically and empirically. Conclusions follow.

I. What Is Corporate Finance?

Starting a new section dedicated to corporate finance, the *Wall Street Journal* defined it very effectively as the “business of financing businesses.” However, Webster’s dictionary defines the verb “to finance” as “to raise or provide funds or capital for.” Why then the adjective *corporate*? What is *corporate* finance about, as opposed to finance in general?

One could argue that the presence of the adjective "corporate" stems from the need to distinguish corporate finance from the other areas of research in finance. However, that would be an excessively academic-centric view of the world. As the *Wall Street Journal's* new section demonstrates, practitioners feel the need to qualify the term "finance" with the adjective "corporate." A more compelling explanation is that the adjective "corporate" helps distinguish corporate finance from other forms of financing, such as real estate finance, which concerns the financing of assets, and from personal/consumer finance, which concerns the financing of an individual or a household. Thus, corporate finance means financing of a firm—not simply an asset, not simply an individual, but that unique combination of assets and individuals that constitutes a firm.

Interestingly, however, the term that prevailed is "corporate finance" as opposed to "enterprise finance" or "firm finance." Although it is true that the main legal vehicle used for almost any economic activity is a corporation, it is not the only one. The choice of this term is probably due to historical reasons. The practice of actively raising capital from a large public of investors for the purpose of undertaking new private ventures started with the spread of the legal concept of a corporation. In fact, during the seventeenth century, early corporations (such as the East India Company) were granted limited liability with a special royal decree for the purpose of facilitating the raising of capital for socially beneficial endeavors that involved too much capital and too much risk to be undertaken by a few wealthy individuals. In spite of some major setbacks, such as the South Sea Bubble, this system proved so successful that after the middle of the nineteenth century England started granting freedom to incorporate to all business enterprises. All other major countries followed promptly. Thus, although financing in some forms goes back as far as Babylonian king Hammurabi (1800 B.C.), it was only after the middle of the nineteenth century that the raising of funds in the *marketplace* became common practice. Not surprisingly, the meaning of the verb "to finance" as "to raise or provide funds or capital for" entered the English dictionary precisely in those years, that is, in 1866.

II. Not Just Semantic

The brief discussion above highlights two important points. First, corporate finance concerns the financing of enterprises, that is, unique combinations of physical and human capital. Second, for historical reasons, the idea of enterprise that became ingrained in corporate finance coincided with the legal notion of corporation. This historical association was not inconsequential. In fact, I claim that even before a direct link between the theory of the firm and corporate finance was established, the field was subconsciously shaped by this identification of the object of study with the legal entity known as a corporation. Let us see how.

A. Limited Liability

The most distinguishing feature of the legal entity called the corporation is limited liability: investors are not personally responsible for corporate liability. As Black and Scholes (1973) pointed out in their seminal article, this feature assimilates a firm's equity to a call option on the firm, having a strike price equal to the face value of the outstanding debt. This option-like nature of equity is behind the asset-substitution effect of Jensen and Meckling (1976) and the underinvestment problem of Myers (1977). These two effects represent the workhorse of the capital structure literature for the last 20 years.¹ Especially before the advent of the literature on control (e.g., Aghion and Bolton (1992)), most papers on capital structure that were not using signaling or taxes were based on some variation of the asset-substitution effect or the underinvestment problem (or both).

B. Corporate Governance

The other important distinguishing feature of a corporation, which is closely linked with its origins, is the large set of laws that regulates its working. Some of these laws are simply default rules, freely disposable by the parties; others, such as the directors' fiduciary duty toward shareholders, are not.² The identification of the enterprise with the corporation, together with the paramount role played by the law in shaping the corporation, has transformed the debate on the governance of enterprise into a debate on corporate law. Although it would be foolish to ignore the importance of corporate law in the entire governance debate, this should not be at the expense of other important factors. If we are willing to accept a broader definition of corporate governance (e.g., Zingales (1998a)), then many other dimensions play an important role. The competitive structure of input and output markets, for instance, is an often-mentioned but rarely studied aspect that might have greater impact on the governance of enterprises than the details of the law. Similarly, the structure and the control of the media industry, which shapes the formation of public opinion and thus the creation of reputation, have been largely ignored in the growing literature on corporate governance comparisons across countries; however, this plays no secondary role.

For example, shareholders' activist Robert Monks succeeded in initiating some major changes at Sears, not by means of the norms of the corporate code (his proxy fight failed miserably) but through the pressure of public opinion. He paid for a full-page announcement in the *Wall Street Journal* where he exposed the identities of Sears' directors, labeling them the "non-

¹In a recent paper, Parrino and Weisbach (1999) have challenged the importance of these two effects from a quantitative point of view.

²In fact, in the legal doctrine it is still highly controversial whether the fiduciary principle is disposable by the parties (see *Everett v. Phillips et al.*, Court of Appeals of New York, June 4, 1942).

performing assets" of Sears (Monks and Minnow (1995)). The embarrassment for the directors was so great that they implemented all the changes proposed by Monks.

Why did the ad work where the proxy fight failed? In the United States, newspapers have a tradition of being accurate and reliable. Readers rely on them to form their opinions. Sears' directors knew it and realized the cost their inactivity had on their own reputation; and this is, according to Fama and Jensen (1983), the most important factor motivating corporate directors. What Fama and Jensen do not emphasize is that this reputation is a function not only of an agent's past behavior but also of the system that transforms this behavior into public information. In Russia, for instance, all the newspapers are controlled by major oligarchs who attack each other on a regular basis with major allegations. In such an environment, how could a director form a reputation for integrity? And why should he care?

C. Dispersed Investors

Although corporations are not necessarily public companies, my brief historical excursus on the origin of corporate finance can explain why large, publicly traded corporations with dispersed investors became its natural object of study. Although numerically these companies represent the exception even in the United States (not to mention in the rest of the world (La Porta, Lopez-de-Silanes, and Shleifer (1999))), they have occupied the center stage both in the theoretical debate and in the empirical analysis, greatly influencing what has and has not been studied.

Theoretically, the emphasis fell on the fact that investors were dispersed and thus unable to coordinate. Hence, the focus is on the distortions produced by the free riding behavior of creditors in financial restructuring (Bulow and Shoven (1978)) or of shareholders in the case of proxy fights or takeovers (Grossman and Hart (1980)). Empirically, this bias toward large companies has led to an excessive concentration of studies on large publicly traded companies, which are certainly the most important ones from a value-weighted point of view but which are also the ones where internal funds are generally abundant and external financing (especially with equity) is a rare event. Thus, most of the empirical effort has been dedicated where we expect finance to matter the least. This explains the great influence the Modigliani and Miller (1958) theorem played, not just as a theoretical starting point but also as a positive description of the world.

Even though this perspective has brought tremendous insights, it has not been without costs on both the theoretical and the empirical fronts. Theoretically, the emphasis on large companies with dispersed investors, for instance, has underemphasized the role that different financing instruments can play to provide investors better risk diversification (for a notable exception see Allen and Gale (1994)). If all companies' stock is held by well-diversified investors, there is little need for additional diversification. Unfortunately, this convenient assumption does not seem to hold in practice. Most companies do have a large shareholder, who is not well diversified.

Even when the financial capital is held by well-diversified investors, the human capital invested in the firm is not well diversified. Thus, a greater attention to these problems is well warranted. Empirically, the emphasis on large companies has led us to ignore (or study less than necessary) the rest of the universe: the young and small firms, who do not have access to public markets.

One could claim that the disproportionate amount of research dedicated to large publicly traded companies is simply an effect of data availability. The access to datasets, such as COMPUSTAT, facilitated this type of research, whereas the difficulty in obtaining information about privately held companies prevented the other type. While there is certainly some truth in this alternative explanation, I think it ignores the extent to which the availability of data is endogenous to the interest of researchers. Consider for instance the long time series of stock prices. Its availability is not a gift from God but the result of the interests of researchers such as James Lorie, who in the early 1960s created the University of Chicago Center for Research on Security Prices, whose data we all love and use today. Similarly, it was only after an intense scrutiny both by researchers and by journalists that data on executive stock options were collected on a regular basis by Standard and Poor.³

III. The Theory-of-the-Firm Foundations of Corporate Finance

The previous section showed that even the term used to define the discipline had great importance in the direction in which the discipline evolved. Much more important, though, have been the theoretical developments in the concept of the firm. In this section, I first argue why the link between the theory of the firm and corporate finance is so tight, and then I draw forth the implications that the most important theories of the firm have for the three main areas of corporate finance:

1. Capital structure, that is, the distortions associated with different ways of financing;
2. corporate governance, that is, the optimal governance structure for this entity; and
3. valuation, that is, the valuation of this entity.

A. Why This Link?

In their seminal article on capital structure, Modigliani and Miller (1958) assume the existence of a firm (actually of two firms in the same “risk class”) with a predetermined payoff. Once we accept this assumption, the irrelevance

³ Of course, disclosure requirements do help in making data available. The Executive COMPUSTAT database was also made possible by the SEC requirements to disclose the details for executive option plans in the proxy statements. Mandatory disclosure requirements themselves, however, are a function of what regulators think is important, which is itself affected by the theoretical model regulators have in mind and thus is ultimately affected by what theory considers important.

of financing choices follows. Thus, as is well known by now, the secret to finding reasons why capital structure does matter lies in a deeper understanding of the content of the firm “black box” and how this content is affected by different choices of financing instruments. An answer to these questions, however, is only possible after we have defined what a firm is and how it operates.

Furthermore, a major role in the capital structure debate is played by the costs of financial distress. Most models assume some costs of liquidating a firm. But where do these costs come from? To understand why value is lost in liquidating a firm, we need to understand why a firm is worth more than the sum of all its components. We must therefore understand what a firm is about and how it adds value with respect to the market. In sum, we must have a theory of the firm.

The link between theory of the firm and corporate governance is even more compelling, and I have already argued for it in Zingales (1998a). The word “governance” implies the exercise of authority. But in a free-market economy, why do we need any form of authority? Isn’t the market responsible for allocating all resources efficiently without the intervention of any authority? In fact, Coase (1937) taught us that using the market has its costs, and firms alleviate these costs by substituting the price mechanism with the exercise of authority. By and large, corporate governance is the study of how this authority is allocated and exercised. But to understand how this authority is allocated and exercised, we need to know why it is needed in the first place. We need, thus, a theory of the firm.

The connection between the theory of the firm and the theory of valuation is probably less apparent, but, precisely because of that, is very important. From an economic point of view, we would like to measure the entire value created by a firm, as the discounted sum of the payoffs generated by the entity called the firm minus the opportunity cost of the inputs used. A theory of valuation, thus, presupposes a definition of what this entity is and an understanding of the relationship between the prices paid for the inputs and their opportunity cost. Unionized workers, for instance, tend to be paid above their opportunity cost. The total value of the firm, thus, should include the rent appropriated by unions. While economically very sensible, this perspective on valuation is clearly different from the one used in most finance textbooks. Finance textbooks identify the value of the firm as the value of all financial claims outstanding. Whether this approach is justified depends on what theory of the firm we espouse. This fact by itself proves the importance of theory of the firm for valuation.

Now that I have, I hope, explained the sources of the connection between the theory of the firm and corporate finance, let us review how this connection plays out in the major theories of the firm.

B. The Firm as a Nexus of Explicit Contracts

The view of the firm prevailing in corporate finance is due to Alchian and Demsetz (1972) and Jensen and Meckling (1976). They define the firm as a nexus of contracts. Sometimes this definition includes only explicit con-

tracts, whereas other definitions include both explicit and implicit contracts. These two notions, though, have very different implications. Thus, I will treat them as two separate views of the firm. In doing so, I will not do justice to many authors, but I will greatly simplify the exposition.

B.1. Implications for Capital Structure

If we consider the firm a simple nexus of explicit contracts, then the firm does not exist as a separate entity but just as a shorthand notation for this set of contracts. By definition, then, the firm cannot be worth more than the sum of the individual contracts that compose it. This obvious conclusion carries enormous consequences for our view of financial distress and, thus, for capital structure. If the value of a firm is simply the sum of its parts, then the prospect of a piecemeal liquidation cannot affect its overall value. Once we accept this, the irrelevance result of Modigliani and Miller (1958) applies straightforwardly.

To be true, one can easily introduce some costs of financial distress in the form of the cost of renegotiating the contracts that belong to the nexus. As a shorthand notation, the nexus saves transaction costs, and thus its dissolution may force costly renegotiations. In fact, Fama (1990) provides a rare example of integration between capital structure and theory of organizations.

Although logically consistent, this approach does not seem to provide a very rich and convincing theory of the costs of financial distress. In the nexus, there are multiple ways to offset changes in one contract or in the environment. Thus, the deadweight loss in renegotiation is bounded above by the cost of renegotiating the simplest (and cheapest) contracts to renegotiate. Consider, for instance, the financial structure of a dividend paying company. The cost of renegotiating the financial structure is bounded above by the cost of adjusting the dividend, which is hardly overwhelming.⁴

B.2. Implications for Corporate Governance

If we accept the view that decision rights should be allocated to the party that can benefit and lose the most from these decisions, then this view of the firm has very sharp implications for the allocation of voting rights. Looking just at explicit contracts, the only residual claim is equity. Thus, shareholders deserve the right to make decisions. Hence, we have the basis for shareholders' supremacy.

To accept this view at face value, one has to take a very legalistic view of contracts. Although *de jure* equity is the only residual contract, *de facto* a firm's decisions influence the payoff of many other members of the nexus, sometimes even to a greater extent than that of equity holders. In fact, Fama and Miller (1972) already point out that bondholders are not fully

⁴ It is certainly possible to argue that the cost of adjusting dividends is very large for a manager, because of the reputational consequences of such a decision (e.g., Myers (2000)). But this is equivalent to admitting that the manager has an implicit contract with the firm, the terms of which depend upon her reputation in the marketplace.

protected by the decisions made by the shareholders. But they dismiss this problem as second order, intuition confirmed by the simulation results of Parrino and Weisbach (1999). But is this true with regard to all the members of the nexus?

If we take the nexus of contracts view literally and accept that contracts specify all the future payoff-relevant contingencies for everybody but the shareholders, then the allocation of decision rights would not be a matter of contention. All the other members of the nexus would be indifferent to the choice made by equity holders because they are contractually protected against any negative consequence. If they were allocated the decision rights, they would happily dispose of them for a cent or could be easily bribed into making the same choices shareholders would make. For control rights to be valuable (as we know they are, see Zingales (1994) and (1995)) and reflect a real conflict of interest, the party in control must be able to make decisions that alter the distribution of payoffs among the members of the nexus. However, this implies that other contracting parties besides equity holders are not fully protected by the explicit contracts, undermining the basic premise of shareholders' supremacy. If many members of the nexus are residual claimants, why are shareholders necessarily the ones affected the most by the firms' decisions? Even if they are, are they the party that benefits the most from the additional protection granted by the control rights?

B.3. Implications for Valuation

If we ignore this problem, however, the nexus of explicit contracts view provides clear insights on how a firm should be valued. Since all other interests are fully protected by contracts, which pay the various contracting parties their opportunity cost, the value created by a firm can be computed by subtracting from the firm's payoff all the contractual payments to the other patrons of the firm. This is equivalent to computing the difference between the payoff received by the only residual claimants (equity holders) and their initial investments. If we ignore the cost of investment, which is sunk, this boils down to the discounted value of future payments to equity. Hence, we have the cash-flow-to-equity approach to firm valuation.

But because within the nexus of explicit contracts the Modigliani and Miller theorem is a reasonable approximation of reality, the cash-flow-to-equity approach can be substituted with the discounted cash-flow analysis (DCF) (see, for example, Kaplan and Ruback (1995)). Thus, the nexus of explicit contracts is really at the heart of most of what we teach to our MBA students.

B.4. Other Implications

Finally, the nexus of explicit contracts view provides clear insights on how a firm should be managed. Because shareholders are the only ones who bear the cost of decisions, the firm should be governed to maximize shareholders' value. At the same time, because any change in total value can be measured by changes in the value of the residual contract, the impact of decisions on

stock prices can be used as a way to evaluate the social consequences of decisions, such as corporate investments, mergers, and so on. Hence, we have the proliferation of event studies.

This approach also has very unequivocal policy implications. It justifies shareholders' supremacy, and it eliminates (at least from a theoretical point of view) the possibility of any inefficiency. Capital structure choices, for instance, are by and large irrelevant, and any government intervention is bound to make things worse.

C. The Firm as a Nexus of Explicit and Implicit Contracts

A variation of the approach outlined above defines the firm as a nexus of explicit and *implicit* contracts (see, e.g., Baker, Gibbons, and Murphy (1999)). This might seem like a minor variation, but in fact it changes the conceptual framework in a dramatic way. First of all, implicit contracts, unlike explicit ones, are not available on demand: they cannot be reproduced simply by signing a photocopy of the original contract. To be credible, they require one party (or both) to have established some reputation over time.⁵ Thus, a firm is not simply the sum of components readily available on the market but rather is a unique combination, which can be worth more or less than the sum of its parts. Consider, for instance, a firm with the reputation of rewarding employees on the basis of their contribution to the firm, regardless of their value in the marketplace. Counting on this reputation, the employees will make investments that are different from those they would have made in the marketplace. If these investments are indeed valuable and could not have been elicited with an explicit contract, the firm's reputation adds value: it represents an organizational asset. On the other hand, if these investments are wasteful, the firm's reputation will destroy value; it represents an organizational liability. The firm, thus, can be worth more or less than the sum of its parts, with the difference being the net of value of organizational assets and liabilities. By using a loaded term, I will call this difference the value of organizational capital.

C.1. Implications for Capital Structure

A theory of the firm able to incorporate the idea of organizational capital can provide a much richer theory of the costs and the benefits of financial distress and thus of the effects of financing on a firm's value. Financial distress can be viewed as the threat that a firm will be liquidated. Such a threat may destroy the effectiveness of the implicit contracts, eliminating the workers' incentive to invest. This might turn out to be detrimental or beneficial to the firm, depending upon the contribution of those investments to the firm's value. What is noteworthy, however, is the fact that a temporary shock (such as financial distress) may have very long-term consequences on the value of the firm. After recovering, a firm may have the

⁵ For an excellent model of how reputation can be acquired over time see Diamond (1989).

same set of objective characteristics as before but may be worth more or less, depending on whether the period of distress has destroyed some of its organizational capital.⁶

Another major difference in this approach compared to its narrower precedent is in the definition of a firm's boundaries. Once we accept the idea that implicit contracts are part of the nexus, the economic definition of the firm starts to differ from its legal counterpart. A firm may have implicit contracts with its suppliers and its customers, but suppliers and customers are not generally considered part of the corporation. This is not necessarily a shortcoming of this approach. At the very least it makes us appreciate that the legal entity undertaking the financing does not coincide with the set of relationships affected by that financing. A neat illustration of this "externality" of the financing choices is a paper by Titman (1984), in which he studies the effects of capital structure choices on the implicit contracts a firm has to service its customers.

If we accept this potential externality, however, we must also accept the possibility that private contracting does not deliver an efficient outcome. Consider, for instance, a company that manufactures a very special product and would maximize its profits if it could credibly promise its customers to service the product for the long term. One possible way to implement such a commitment would be to maintain a very high credit rating, which would make it very unlikely that the firm would be liquidated any time soon. Depending on the temporal pattern of sales, however, this commitment might not be feasible, which would reduce the value of the company. In such a case, a regulation that forces all the companies in that sector to maintain a high credit rating to operate might help in making this commitment credible.⁷

In fact, the very need for implicit contracts suggests the existence of some inefficiency in explicit contracting, which can be used to claim the benefits of government intervention. But this inefficiency only constitutes what Coase (1990) calls "prima facie evidence" in favor of regulation. To ascertain whether such intervention is really justified, we need a better sense of what imperfections implicit contracts are trying to address and to what extent regulation can improve upon a private solution. In sum, we need to have a more fully developed theory of the firm along these lines.

C.2. Implications for Corporate Governance

If taken seriously, this theory of the firm also has governance implications that are at the opposite end of the spectrum with respect to the traditional nexus of contracts view. In fact, once we recognize the existence of implicit contracts, then there are other residual claimants besides equity holders

⁶ For an attempt in this sense, see Novaes and Zingales (1993).

⁷ Of course, one could argue that the initial shareholders could devise a similar contract with the product customers up front. This contract, however, will face two difficulties. First, it will be extremely difficult to renegotiate in case such a renegotiation is needed. Second, it would require a very expensive bond up front.

who may need to be protected (the famous stakeholders, often mentioned in the public policy debate). It then becomes unclear whether control should reside in the hands of shareholders, because the pursuit of shareholders' value maximization may lead to inefficient actions, such as the breach of valuable implicit contracts, as described by Shleifer and Summers (1988).

C.3. Implications for Valuation

Finally, the presence of implicit contracts in the nexus also raises two important issues for valuation. First, the members of the nexus can be paid above or below their opportunity cost in different moments of their relationship. This creates a wedge between market prices and the opportunity cost of inputs, which makes it impossible to identify the value created by a firm with the payoff accruing to equity holders. As a result, stock price changes are not reliable indicators of welfare changes even when the market is perfectly efficient. This possibility warns us against drawing any welfare conclusion from event study analysis.

Second, this view of the firm recognizes the existence of hidden organizational assets (and liabilities). Understanding and valuing them is a major challenge for future research.

C.4. Summary

In sum, it is interesting to highlight how a slight variation in the definition of the firm can trigger major changes in the implications. Once implicit contracts are added to the nexus, one can easily justify government regulation of financing choices, undermine many results obtained in the event study literature, and rationalize other stakeholders' theories of corporate governance. This alternative paradigm, however, is not very well developed, possibly because of the difficulties in working with implicit contracts. Most of the implications, thus, remain to be worked out formally.

D. The Firm as a Collection of Growth Options

Myers (1977) defines the firm as a collection of assets in place and growth opportunities. This is not a full-fledged theory of the firm, because it does not provide explanations of what the glue is that keeps the growth opportunities attached to the assets in place and of how the inner workings of this entity differ from the workings of the marketplace. Nevertheless, this approach has been extremely important in shaping the capital structure debate and has had profound and long-lasting effects on the theory of valuation.

D.1. Implications for Capital Structure

The interesting insight that this view brings to the financial structure debate is the importance of the interaction between assets in place and growth opportunities. In both of Myers's greatest contributions (Myers (1977) and Myers and Majluf (1984)), it is the structure of the existing assets and lia-

bilities that impedes the efficient exploitation of growth opportunities, when these need to be financed with external equity. Informational asymmetries on the value of the assets in place between insiders and outsiders are responsible for the friction in Myers and Majluf (1984). If a new project requires equity financing and the manager has private information that makes her believe the market is undervaluing the assets in place, she will prefer passing up the valuable growth option rather than diluting the value of the existing shareholders. On the other hand, it is the excessive amount of debt raised to finance the existing assets that makes it very costly for shareholders to raise new equity because the proceeds of the equity issue go to insure the existing debt in case of default, increasing its value at the expense of equity.

Both these papers provide a clear view of the costs of financial distress: underinvestment in valuable projects. But what is the origin of the problem? In both cases, the problem arises because the growth option cannot be easily detached from the existing firm to become a separate company. Thus, the essence of the problem is the inextricable connection between assets in place and growth opportunities. But understanding this connection is equivalent to having a theory of the firm able to explain why a firm is worth more than the sum of its parts. I do not think we are there yet.

D.2. Implications for Corporate Governance

I am not aware of any formal development of the consequences of this approach for corporate governance. This is simply a reflection of the lack of a full theory explaining why assets in place and growth opportunities are linked together and how the right to exploit these opportunities is allocated. Once such a theory is in place, the role of a governance system will be to ensure that the power to make decisions is allocated to the people with the best opportunities.

D.3. Implications for Valuation

Finally, Myers's (1977) approach has revolutionized valuation techniques. The standard discounted cash-flow model values a firm on the basis of the cash flow that a firm will generate, net of the cost of making the future investments. Future investments, however, are discretionary and will be made only when and if they will generate a positive expected net present value. Future investments, thus, are options that the firm can exercise at will, and as such they need to be valued. The real option approach to investments and valuation is only now spreading outside academia and becoming a common, although not yet standard, technique in the industry.

Although the similarity to financial options has been extremely fruitful in enhancing our understanding and in providing analytical tools to improve valuation techniques, it has led the literature to ignore some important differences between the two. Financial options are well-specified contracts with a clear owner and a defined payoff. Except for exploration rights, real options rarely share these features. Many real-world growth options are not

clearly allocated to one owner, and their payoff is highly dependent upon the way the option is exercised; that is, it is endogenously determined, rather than exogenously specified, as in the case of financial options.

Consider, for instance, a food company composed of several existing brands and numerous options to introduce new brands through its commercial channels. The ownership of the existing brands is well defined and well protected: anybody else who tries to market a product under an existing brand name can be stopped. The ownership of the option is less protected. To start with, the idea of a new product will come to an employee who has the choice whether to pursue it within the company or on her own. Although the law prevents flagrant exploitations of corporate opportunities, there is usually ample leeway to do so without triggering any effective legal consequence. In fact, Bhide (2000) reports that 71 percent of the firms included in the Inc 500 (a list of young, fast-growing companies) were founded by people who replicated or modified an idea encountered in their previous employment.

Even if the laws were able to effectively prevent employees from exploiting a firm's real options, the firm would rarely enjoy a monopoly in the exploitation of its growth opportunities. Unlike financial options, in real options there are potential competitors who can seize the opportunity or take advantage of a delay in exercise to appropriate most of its net present value. Finally, even in the absence of competitors in the product market, the firm might face the potential holdup by crucial suppliers, who although unable to exercise the option on their own, are happy to exploit their monopolistic position to capture some (or all) of the value embedded in these options. One of the most famous examples, described in Klein, Crawford, and Alchian (1978) (see also Williamson (1985)), is that of Fisher Body holding up General Motors when an unexpected surge in the demand for cars forced GM to purchase its car bodies from Fisher Body. Another example is sequels of successful movies. The right to make a sequel can be seen as an option that can be exercised by the studio that produced the original movie. Except for animated movies, where the studio owns the entire value, in regular movies the crucial actors "own" a claim on part of the net present value of that option. Because it would be hard to have a Godfather movie without Al Pacino or a Terminator movie without Schwarzenegger, these actors effectively own a stake in the sequel. Consistently, the cost of producing a sequel is in general 20 percent higher than that of the original movie (Luherman and Teichner (1992)).

As I will argue in Section V, we really need a theory of the firm to lay down the factors underlying a firm's ability to capture the value of new growth options.

E. The Firm as a Collection of Assets

One of the major shortcomings of the traditional nexus of contracts view is that it is unable to explain what is so special about mergers. If the firm is simply a collection of contracts, the results achieved through a merger could be more simply obtained by writing a contract between two separate firms.

Why, then, do firms spend billions to merge? The property rights literature developed by Grossman and Hart (1986) and Hart and Moore (1990) addresses this question. The starting point is that contracts are necessarily incomplete; and, as a result, there is a need to allocate the right to decide in the events not specified by the initial contract. This decision right, which Grossman and Hart (1986) associate with the ownership of an asset, affects the distribution of the ex post surplus created by an enterprise and, thus, the incentives to generate this surplus. By changing the allocation of ownership, a merger changes these incentives in a way that no contract could. In the property rights literature, thus, what makes a difference is how ownership is allocated. Consistent with this view, the firm is defined as a collection of assets that are jointly owned.

The concept of residual rights of control, introduced by Grossman and Hart (1986), is a fundamental contribution, which has shaped indelibly the whole subsequent discussion. Another appealing feature of their definition of the firm is that it corresponds very closely with the legal definition. A potential shortcoming is the identification of control with ownership, which Hart and Moore (1990) further specialize in the right to withdraw the asset. As Kay (1996, 111) puts it very effectively: "if we asked a visitor from another planet to guess who were the owners of a firm [on the basis of this definition] by observing behaviour rather than by reading text books in law or economics, there can be little doubt that he would point to the company's senior managers." Not surprisingly this line of research has found it extremely difficult to deal with the separation between ownership and control.

E.1. Implications for Capital Structure

The identification of both ownership and control with the right to withdraw the asset has led to very elegant theories of debt (Aghion and Bolton (1992), Hart and Moore (1989), Bolton and Scharfstein (1990), Hart and Moore (1994)), but it has failed to deliver a convincing model of outside equity. The few attempts made (Fluck (1998), Myers (2000)) model outside equity as a sort of debt with infinite face value, which always has the ability to intervene and subtract the asset from the managers. This characterization seems to underplay the diversity of equity from debt and overplay the power to intervene of outside equity.

A possible solution within this framework is to find a role for control that is different from that of ownership. Rajan and Zingales (1998), for instance, identify the separate role of control in the regulation of access to the asset itself. This can be a valuable instrument if the asset requires specific investments by the employees. One can then distinguish between 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 the access to the asset before specific investments are made.

Another shortcoming of this definition is that it focuses on the firm after all the human capital specific investments have been done. At this stage, the firm is simply the sum of its parts (assets), and the only reason why liqui-

dation can produce a loss in value is if the market for individual assets is less liquid than the market for the whole firm (e.g., Shleifer and Vishny (1992)).

E.2. Implications for Corporate Governance

The difficulty of this approach in dealing with the separation between ownership and control has also impinged on its ability to deliver implications for corporate governance. In fact, there is some schizophrenia between the message contained in the original contributions and the way these contributions have been interpreted. For example, there is a clear Marxian flavor in the role played by ownership in Grossman and Hart (1986) and Hart and Moore (1990). The residual rights of control allow the owner to extract the surplus out of the worker. The key workers, thus, should control the assets they work for to prevent this “exploitation.” Unlike in Marx, however, this reallocation of ownership is not motivated by distributional concerns but by efficiency concerns. If workers expect to be exploited, they will not make valuable investments. Although I am not aware of any corporate governance paper arguing along these lines, I think this would be a reasonable literal interpretation of the property rights view of the firm.

Interestingly, this is not the way this view has been used. Quite to the contrary, it has been reinterpreted to support the shareholders’ value maximization paradigm. The best example is Shleifer and Vishny’s (1997) corporate governance survey. They recognize that contracts are incomplete and that the allocation of the residual rights of control shapes the ex post distribution of surplus. But they argue that insiders have a better bargaining position as a result of their ability to quit. Outside financiers—they conclude—should be protected against expropriation through the residual right of control. The greater need for protection, however, only suggests that the suppliers of finance need some form of protection; it does not necessarily imply that they should be protected via the residual rights of control rather than via other contractual mechanisms. This is not a foregone conclusion, because Rajan and Zingales (2000b) show that it might be efficient to allocate formal control rights to the party who has a lot of *de facto* power, as in the case of workers who can easily leave.

E.3. Implications for Valuation

Some potential contradictions also emerge when we try to move from the property rights view of the firm to a theory of valuation. The theory makes it quite clear that there is not a fixed or exogenously specified division between cash flow going to financial claimholders and cash flow going to insiders. Therefore, it would be natural to define the value of the firm as the total surplus produced by the firm, which is then divided between these two groups. However, by defining the firm as a collection of assets, the property rights view excludes the insiders’ human capital. As a result, the value of the firm gets identified with the value that can be appropriated by the owner of the assets, thanks to their control of the assets. The amount of surplus ap-

propriated by insiders, beyond what they are entitled to on the basis of their equity stake, is then classified as *private benefits*. The concept of private benefits has been introduced precisely to reconcile this tension between a definition of a firm's value based only on the amount appropriated by financial claimants and a model where the division between insiders and outsiders is endogenous. On the one hand, the introduction of private benefits has been an important step toward recognizing that there is an appropriability problem and that financial structure and governance structure affect who is able to appropriate what.⁸ On the other hand, it has often confused the problem, as is testified by the ambiguous use of private benefits in the welfare analysis. Although some papers explicitly account for them (e.g., Aghion and Bolton (1992)), others ignore them (e.g., Grossman and Hart (1980)) on the basis that their value will be dissipated in rent seeking.

F. What about Signaling?

By focusing on the link between financial structure and the theory of the firm, I left out an important branch of the capital structure literature: the one based on signaling. By assuming that managers have an incentive to reveal their superior information to the capital market via the choice of different financing instruments, this literature has derived interesting implications for capital structure (Ross (1977)), dividends (Bhattacharya (1979); John and Williams (1985), etc.). In fact, the richness of this literature might appear as a contradiction to my claim that any work in corporate finance is rooted in an underlying theory of the firm, because most of this work seems to be independent of any notion of the firm.

To the contrary, this is the exception that proves the rule. This literature has gone out of fashion not for lack of empirical support but for its difficulty in addressing questions such as: Where do the (assumed) incentives to signal come from? Why have certain devices and not others been chosen? To answer these questions, we need a theory of the firm in the background. And when such a theory is provided, the effects of signaling may disappear. For example, within a nexus of contract framework Dybvig and Zender (1991) show that if we endogenize the contract shaping the incentive to signal, then the irrelevance proposition still holds. Thus, even the signaling literature is strictly dependent upon an underlying theory of the firm.

IV. The Nature of the Firm Has Changed⁹

In spite of some criticisms I have voiced, the existing theories have been quite successful in providing an intellectual backbone for the principles of corporate finance we all know and teach. It is less clear, however, whether this model will be helpful very much longer. In fact, the very model of a

⁸ The most extreme case of appropriability is when insiders can dilute the value of the company to their own benefit (see Zingales (1994), (1995), and (1998b)).

⁹ This section is based on Rajan and Zingales (2000a).

corporation that has inspired the debate on theory of the firm and, indirectly, informed the theoretical debate in corporate finance is changing, and many of the conclusions we took for granted need to be reexamined.

The case of the British advertising agency Saatchi and Saatchi, described in Rajan and Zingales (2000a), well illustrates the clash between the principles derived on the basis of yesterday's model and today's reality. In 1994, Maurice Saatchi, chairman of Saatchi and Saatchi, proposed for himself a generous option package. The U.S. fund managers, who controlled 30 percent of the shares, became furious. The stock had underperformed for several years, and the last thing they wanted was to reward the chairman. This act of managerial self-interest needed to be punished with a serious shareholders' initiative, and so the shareholders voted down the proposal at the general shareholders' meeting. This opposition led to the departure of Maurice Saatchi, quickly followed by the resignation of several key senior executives. These executives, together with the Saatchi brothers, started a rival agency (M and C Saatchi) that in a short period of time captured some of the most important accounts of the original Saatchi and Saatchi. The original firm, which later changed its name to Cordiant, was severely damaged.

In hindsight, the mistake the U.S. fund managers made was to treat Saatchi and Saatchi as a traditional company with clear boundaries defined by its assets. Because they had ownership (thanks to their 30 percent holding of the votes) they may have thought they controlled the firm. Instead, much of the firm broke off as they attempted to exercise their traditional ownership rights.

Interestingly, in 1994 a firm like Saatchi and Saatchi, with few physical assets and a lot of human capital, could have been considered an exception. Not any more. The wave of initial public offerings of purely human capital firms, such as consultant firms, and even technology firms whose main assets are the key employees, is changing the very nature of the firm. Employees are not merely automata in charge of operating valuable assets but valuable assets themselves, operating with commodity-like physical assets. The changing nature of the firm forces us to reexamine much of what we take for granted in corporate finance. What precisely is the entity that is being financed, governed, and valued?

A. *The Prototypical Traditional Firm*

There are four main features of the traditional firm that are relevant from our point of view. First, the traditional firm, which according to Chandler (1990) emerged during the second industrial revolution to exploit economies of scale and scope, was very asset intensive and highly vertically integrated. It was asset intensive because the first companies in each sector that could exploit the economies of scale and scope gained a formidable advantage vis-à-vis new entrants (Chandler (1990)). At the same time, it was very highly vertically integrated, because the need to ensure the right level of throughput, at a time when the market for intermediate goods was just developing, forced companies to take direct control of their suppliers and distribution

systems (Chandler (1977)). As a result, the realm of transactions governed by power rather than by prices tended to coincide with the legal boundaries of the corporation.

Second, the traditional firm had a high degree of control over its employees. Vertical integration reduced the size of the market for intermediate products. The scarcity of competitors, both in the intermediate and in the output market, implied a thin outside labor market able to use (and pay for) the skills that employees acquired on the job. Through its control of the firm's assets, the headquarters effectively controlled the main source of employment open to its specialized employees, giving to top management enormous power.

Third, the size and the asset intensity of the traditional firm required more investment and more risk taking than were within the capacity of the management. The control conferred by the ownership of crucial assets, however, made outside ownership feasible. Therefore, the traditional firm came to be owned by disperse investors.

Finally, the concentration of power at the top of the organizational pyramid, together with the separation between ownership and control, made the agency problem between top managers and shareholders the problem. Both capital structure and corporate governance became singly focused on this dimension. The objective became to maximize the protection of outside investors, and the means was by reducing or removing all the obstacles to shareholders' control. Thus, transparency, accountability of directors, contestability of corporate control, and managerial compensation aligned with shareholder wealth maximization became steps in this battle.

B. The New Firm

In the last decade we have witnessed three major changes in the balance of power within firms. First, physical assets, which used to be the major source of rents, have become less unique and are not commanding large rents anymore. Improvements in capital markets, which have made it easier to finance expensive assets, have certainly contributed to this change, as has the drop in communication costs, which reduced the importance of expensive distribution channels, which favors the access to the market of newly formed companies.

Second, increased competition at the worldwide level has increased the demand for process innovation and quality improvement, which can only be generated by talented employees. Thus, the quest for more innovation increases the importance of human capital.

At the very time human capital has become more important, firms' grip on it weakened for two reasons. First, the easier access to financing has increased employees' outside options.

Second, the opening up of world trade created the space for many independent suppliers. This generated many alternative employment opportunities, making employees' human capital less specific to their current employer. At the same time, the increased competition at the intermediate goods level

has also triggered the breakdown of the vertically integrated firm. The rising of competitors against which to benchmark performance exposed the real costs of cross-subsidies, putting pressure on firms to break up.

Summarizing, new firms tend to be non-vertically integrated, human-capital-intensive organizations that operate in a highly competitive environment. In addition, the exercise of authority by the headquarters is severely limited by the ability of employees to quit, taking with them their human capital or (as in the case of Saatchi and Saatchi) part of the firm.

V. A New Agenda

The changing nature of the firm forces us to abandon the illusion that firms' boundaries are clear cut and remain unchanged when we change the capital structure or the governance structure. This was a sensible assumption for a GM type of organization—very physical-asset intensive, where the boundaries were clearly set in advance and not easy to alter—but is not very sensible for today's organizations, which are mostly human capital intensive. In these organizations internal distributional conflicts can change the boundaries of the organization itself, as the Saatchi and Saatchi example shows.

Once we admit that, the need for a theory-of-the-firm foundation of corporate finance becomes even more compelling. Although in Section III I argued that such need has always existed, in the past, the legal notion of a corporation captured closely enough the economic boundaries of the firm that such an explicit link was less necessary. When, for instance, assets in place assigned to their owner a *de facto* exclusive right to pursue further investment opportunities, there was not much need for a theory that explained who can appropriate those options to value a company. When different capital or governance structures did not have much of an effect on who appropriated those options, the corporate finance implications could be safely worked out without taking a position on why growth opportunities were attached to the assets in place.¹⁰ When firms' assets represented the main source of value and control, it was perfectly legitimate to approximate a firm with its physical assets. As a result, most corporate finance theory was successfully developed even if the fundamental questions in the theory of the firm were not completely resolved.

But now that the boundaries of the firms have become fuzzy, employees' human capital has become the most valuable asset, and growth options are up for grabs, answering these questions becomes unavoidable. After all, what defines *corporate* finance is its focus on the challenges raised by financing the unique combination of assets and people that we call firms. Understanding what makes this combination unique is a fundamental step we cannot postpone any longer.

¹⁰ To be accurate, Myers (1977) contains a very interesting discussion on this very issue. Many subsequent developments and applications, however, take the connection between assets in place and growth opportunities for granted.

A. What We Need

The defining characteristic of a firm is that it substitutes authority for the price mechanism in determining how decisions are made (Coase (1937)). The fundamental question a theory of the firm should address, then, is how an organization succeeds in acquiring power that differs from “ordinary market contracting between any two people” Alchian and Demsetz (1972, 777). Only by understanding the source of this power can we hope to explain where the influence of this power ends (i.e., the boundaries of the firm) and how this power operates within the firm’s boundaries.

In the case of physical asset-intensive firms, an answer to this question is already contained in Grossman and Hart (1986) and Hart and Moore (1990). The residual rights of control over the assets attributes to the owner a power that she does not have in ordinary market contracting. The new challenge is to explain what happens when there are no physical assets involved or when these assets are simple commodities and, hence, easily replaceable. This is tantamount to explaining how a third party can have control over human capital in a world where labor has been liberated not only from slavery but also from the modern form of indenture represented by the uniqueness of the assets that labor specialized to. In other words, what is the power of a shareholder in a firm like Saatchi and Saatchi? Answering this question will enable us to explain the role of outside equity in human capital-intensive firms and thus provide a new perspective to the capital structure debate.

The second question a theory of the firm should address is how this power is maintained and enhanced and how it is lost. An answer to this question can explain what the factors are that underlie a firm’s ability to grow or its failure to do so, even in the face of valuable growth opportunities. It will also be a critical step toward an understanding of the costs of financial distress. In fact, as I mentioned in Section III, one of the shortcomings of the existing theories of the firm is their inability to explain why temporary shocks to liquidity can have permanent, and sometimes irreversible, consequences on a firm’s value. If we can show that liquidity shocks alter in an irreversible way the factors responsible for growth, then we would have a new theory of the costs of financial distress and consequently a new capital structure theory.

The third fundamental question a theory of the firm should address is how this authority-based system operates in a way different from ordinary market contracting. Not only would an answer to this question enable us to understand theoretically the effects of mergers and spin-offs, but it also would bring new life to the debate on the costs and benefits of corporate diversification.

Understanding the internal workings of firms will also enable us to derive some policy rules on how a firm ought to be governed. In the simple nexus of contract world, the maximization of a firm’s value corresponded, with few exceptions, to the maximization of shareholders’ value. Consequently, the traditional precepts of corporate governance were all aimed at empowering shareholders by reducing the cost of collective action. In the current envi-

ronment, where human capital is crucial and contracts are highly incomplete, the primary goal of a corporate governance system should be to protect the integrity of the firm, and new precepts need to be worked out.

The fourth and final question a theory of the firm should address is how the surplus generated by the firm is allocated among its members. This is not a trivial task, because by definition, resources inside the firm are not allocated on the basis of prices. Such an understanding will help us derive a new approach to valuing firms that is consistent with the new nature of firms. In fact, the enormous success of the nexus of contracts view of the firm is due, at least in part, to the very practical implications it provides for valuing firms.

The growing importance of human capital and its greater ability to capture rents make this approach no longer tenable. In assessing the value generated by a firm we have to consider the surplus captured by other stakeholders. A more reasonable approach to valuation, thus, would be to discount the total value added generated by the firm. To move from here to the equity value, however, we need a theory of how the surplus is divided among different claimholders, be they financial claimholders (equity holders and debt holders) or nonfinancial ones, such as employees, key customers, and suppliers. Understanding the internal allocation of surplus, thus, is a necessary step toward a theory of valuation of the new firm.

B. An Attempt

Thus far, the discussion has been rather abstract, as is necessarily the case when one tries to illustrate the features of a theory that does not yet exist. To make this discussion a little more concrete, I will describe some attempts Rajan and I have made to develop such a theory. Far from being the solution, this is just meant to be an example of what can be done in this direction, an example that allows me to describe more concretely what the consequences of this approach could be.

In Rajan and Zingales (1999) what distinguishes the firm from the market is the web of specific investments built around a critical resource. To acquire power over human capital, an entrepreneur needs control (not ownership) over some valuable resource, which need not be a physical or alienable asset but can even be her own human capital. By controlling a critical resource an entrepreneur can influence the accumulation of specific investments so as to build complementarities between the person the entrepreneur seeks to have power over and her critical resource. Once the complementarity exists, the specialized person may obey orders from the entrepreneur for fear that disobedience would jeopardize the joint value they can create together. More generally, whereas ownership legally links an inanimate asset to a firm, complementarities economically link some person that cannot be owned to the firm.

But why would anyone choose to specialize when he knows that this will make him dependent on the critical resource and, thus, at the mercy of who controls it? The secret is to create a situation where employees know that

their rewards will be greater if they make firm-specific investments. The enterprise does this by giving key employees or units privileged *access* to the enterprise or its critical resources, so that they have power if they specialize (see Rajan and Zingales (1998)). Through regulated access an entrepreneur can build a nexus of specific investments able to produce more than any competing entrepreneur with a similar resource starting from scratch. This is because the web of specific investments built over time cannot be reproduced instantaneously. This web could also be viewed as a firm's production capabilities (Wernerfelt (1984)) or its organizational capital (Klein (1988)).

In this framework, power is maintained and increased by having more and more people specializing; thus at some point the critical resource becomes the web of specific investment itself. To maintain the incentive to specialize, however, employees have to be granted a certain level of rents in the future. Because the possibility of renegotiation jeopardizes long-term commitments, growth prospects are crucial in making the promise of future rents credible. If growth prospects diminish (as in the case of a temporary liquidity shock), then employees may lose the incentives to specialize, reducing the value of organizational capital (Rajan and Zingales, 1993). This might be a new way to formalize the idea of the costs of financial distress as a temporary shock causing a permanent loss in organizational capital. Note that a similar loss will be generated if the firm were to reduce its growth prospects for non-financial reasons. This might explain why it is so difficult to empirically disentangle economic from financial distress.

A fundamental difference between the working of the firm and the working of markets is the protection granted to property rights in the two environments. Employees are generally not granted property rights inside the firm, whereas this occurs in markets. This lack of internal property rights makes resources more easily transferable within firms than across firms. In fact, one of the benefits of corporate diversification is the ability to transfer resources to divisions with the best opportunities (the so-called winner picking, see Stein (1997)). The flip side is that divisions might distort their investments to protect their own turf. The fear of inefficient protective investments may induce the headquarters of a diversified firm to misallocate some funds at the margin (relative to the first best) to prevent greater average investment distortions. Rajan, Servaes, and Zingales (2000) argue that this is more likely to occur when divisions are very different in their opportunities, and by observing the allocation of funds of diversified firms in the United States, they find evidence consistent with this claim. These results suggest that the costs of corporate diversification tend to outweigh the benefits when divisions have diverse opportunities and resources. This could explain the success of General Electric under Jack Welch's reign. Welch's principle to divest any division that is not the leader in its market segment can be interpreted as a rule meant to reduce the level of diversity inside General Electric. This is an example of how a theory of the firm can be used to shed new light on a highly debated topic (corporate diversification), provide new testable implications, and even rationalize popular management principles.

B.1. Implications for Capital Structure

As I discussed above, a model of how power is acquired and maintained can be used to derive a theory of outside equity. Although this has not been done in this context yet, it is easy to sketch how it could be done. In Rajan and Zingales (1999) outside ownership can be supported only when the web of specific investments around the critical resource is strong enough that the ownership of the critical resource, even without daily control, confers some power. For this to happen, though, an organization should accordingly be designed and modeled over time. This, in fact, may be the major service venture capitalists provide emerging firms (Rajan and Zingales 2000a). They ensure that the management of a firm is professionalized so that it is not too dependent upon the entrepreneur or any specific professional manager. In this way, they make the firm easier to finance, initially by themselves and eventually by dispersed outsiders (for some evidence in this sense see Hellmann and Puri (1999)). Another way to ensure more power to outsiders is to provide a firm with an internal bureaucracy that generates verifiable information on employees' performance (Novaes and Zingales (1998)).

The above argument, however, only explains how outside equity is feasible, not why it exists and who should own it. Of course, one can appeal to the usual arguments that an entrepreneur needs to raise funds and diversify risk. But this approach could deliver some interesting new roles for capital structure: not as financing instruments but as a glue to preserve the rents of an organization from being dissipated by competition among different stakeholders. A recent example of this phenomenon is provided by Autodaq.com, one of the many e-commerce firms that sprouted in the last couple of years. It specializes in the resale of used cars from car rental and leasing companies to car dealers. One of the major problems for Autodaq.com, as for most Internet business, is how to make sure that the efficiency gains generated by its Internet-based intermediation system are not fully passed to the buyers—in other words, how to lock in crucial relationships that enable the firm to appropriate at least a fraction of the benefits. The strategy followed by Autodaq.com is to make the crucial suppliers of used cars shareholders in its company. As shareholders, they will be cointerested in the success of Autodaq.com, making it unlikely they will supply their cars to new entrants. This is one of the many examples where equity is used not to raise funds but to protect the long-term viability of the firm.

B.2. Implications for Corporate Governance

In the traditional type of firm power was clearly allocated to the person controlling the asset; thus the corporate governance problem centered on how to prevent the person holding this power from abusing it. This explains the focus on the agency costs between providers of financial capital and managers and/or between large shareholder and small shareholders.

The deintegration of the firm and the growing importance of human capital are changing the terms of the problem. Power and rents are not concentrated at the top of a steep pyramid; they are sprinkled throughout the

organization, even outside the legal boundaries of the firm, as is the case for crucial independent suppliers. Now that power is diffused, the major corporate governance problem becomes how to prevent conflicts among stakeholders from paralyzing or destroying the firm. In fact, traditionally one of the great advantages of the corporate structure was that it concentrated all the control rights to a group with very homogeneous interest: the shareholders (Hansmann (1996)).

The recent changes in the nature of the firm described above are undermining this advantage. Some patrons become equity holders (as in the example of Autodaq.com), and others maintain enormous power even without controlling votes (as with Maurice Saatchi in Saatchi and Saatchi). These changes make new firms more similar to cooperatives, rather than to traditional corporations.¹¹ Thus, corporate governance has a lot to learn from a serious analysis of cooperative governance along the lines of the one conducted by Hansmann (1996).

At the same time, the increased power of human capital forces corporate governance to confront the fundamental issue of how we should allocate de jure control rights, when there are multiple sources of de facto control rights (Rajan and Zingales (2000a)). This analysis can deliver novel implications both in terms of the optimal allocation of ownership and in terms of the optimal capital structure. For example, Rajan and Zingales (2000b) argue that de jure control rights should be given to the party who has de facto power to minimize the amount of resources wasted in fighting. A similar point is made by Welch (1997) in the context of the seniority structure of debt. He argues that banks, which are better at fighting in court, should hold senior debt, because this will minimize the resources wasted in legal battles.

This fragmentation of power also presents serious risks for the survival of organizational capital. Because no party is the sole residual claimant and no party has absolute control, nobody fully internalizes the preservation of organizational capital. One might argue that this is less of a concern today, because the deintegration of firms has made the organization less specific and thus the organizational capital more portable across firms. But if it is still an issue, we should devise some corporate governance guidelines aimed at protecting the organizational capital against the centrifugal (and destructive) forces represented by the interest of each stakeholder.

One form this disintegration of organizational capital can take is the failure to exploit growth opportunities. If we identify a firm with the nexus of specific investments, the link between assets in place and growth opportunities is represented by their common reliance on the same nexus of specific human capital investments. Whether these opportunities will be exploited, though, depends upon who has the formal right to exercise them in the name of the firm, who has the opportunity to exercise them outside the realm of the firm, and who will benefit most in either case.

¹¹ Hansmann (1996) differentiates between cooperatives and corporations on the basis that in the former capital is supplied by patrons, whereas in the latter it is not.

In some sense, this is the old Myers underinvestment problem in a broader context. In Myers (1977) the failure to exploit firm growth opportunities is due to the fact that the exogenously specified payoff of the party in control (equity holders) does not compensate it enough for its cost of financing the investment. In this broader context, where the payoff of all the parties involved is endogenously determined through bargaining, growth opportunities will not be exploited when they tilt the power distribution against the group who has to pay the “exercise price” of the option. Consider for instance the failure of IBM to market its personal computers through mail order as discussed in Hamel and Prahalad (1994). The real reason for this decision (or lack of it), they argue, is the enormous clout the salespeople and the existing distribution network had at the top of IBM. Although such a move was efficient, it would have had an extremely negative impact on the power of the existing sales network—which successfully fought this choice for a long time—at the expense of IBM’s performance as a whole. A similar failure recently occurred in Bertelsmann, a German conglomerate, with divisions both in publishing and new media. For both of these divisions, the development of book sales through the Internet provided a wonderful opportunity. Yet both ignored the opportunity, for fear that once the unit was developed (and most of the costs incurred) it could be reassigned to another division, subtracting its potential benefits.¹²

This discussion suggests that the primary role of a governance system is to ensure an alignment between the ability to capture the opportunities and the reward stemming from them. A failure to do so will result in a *governance overhang*, not unlike the debt overhang described by Myers (1977). Unfortunately, however, preventing this governance overhang is not as simple as maintaining a low level of debt. In fact, deriving such governance guidelines is a major task for future research.

B.3. Implications for Valuation

To illustrate the opportunities of applying these ideas to valuation, consider the case of Technical Data Corporation (TDC), an early provider of financial information online, described in Sahlman (1983). At the time of the case, which centers on TDC valuation, the firm had already established itself as a reliable bond data provider with a good clientele. On top of this, there were plans to expand the business to include providing on line information about equity. The central question of the case is whether the price should incorporate TDC’s future growth opportunities. These options clearly do not belong to TDC in a legal sense, but this is not the point. Whom do they belong to in an economic sense (i.e., who is going to reap the proceeds from them)? Do they belong to TDC as a whole; to Jeff Parker, its founder, chairman, and large shareholder; or to somebody else?

If we accept the idea that the firm is a nexus of specific investments, what keeps a firm together is the strong complementarity between human and physical capital. Thus, an option “belongs” to a firm if it is highly comple-

¹² See *The Economist*, November 21, 1998, survey p. 10.

mentary with the physical and human capital that constitutes the firm. In the case of TDC, there was a strong complementarity in the distribution channels: the same computer terminals disseminating information about bonds could be used to disseminate information about equity. Similarly, there was some complementarity between the software used in the two cases and thus in the human capital of TDC programmers. There was not, instead, a strong complementarity between Jeff Parker and the growth option, because his area of expertise was the bond market. Thus, in this case it is fairly simple to see that these growth options belonged to TDC as a company (and thus also to the outside shareholders), not to Jeff Parker himself.¹³

Unfortunately, not all the cases are as simple as TDC, and I think we are still very far from having a coherent new approach to valuation. What I hope these examples can convey is the far-reaching implications that a new theory of the firm could have on the way we value companies. For this to happen, however, we also need a great deal of empirical work substantiating some of the crucial steps required to move from an abstract theory to a very concrete valuation.

First, there is a pressing need to understand what factors determine the ability of firms to capture new growth opportunities. The theory focuses on the degree of complementarity. But how can we measure it? And what is the predictive power of such an idea? An answer to these questions will be crucial to deriving the value of the entire enterprise.

Second, we need to know how the surplus generated by a firm is allocated among its members. The dominance of the nexus of contracts paradigm has blinded empirical research on distributional issues inside the firm. With few notable exceptions, such as Rose's (1987) analysis of the distribution of regulatory rents between unions and equity holders, the empirical literature has just assumed shareholders are residual claimants and has ignored how the surplus is split and what factors drive these splits. Such studies would be of tremendous help in moving from the value of the entire enterprise to a value of financial claims.

B.4. Toward a Theory of Entrepreneurship

Understanding the factors underlying a firm's ability to exploit growth opportunities will not only affect the way we value firms but also create a fundamental building block toward a theory of entrepreneurship. Entrepreneurship is the process by which new firms are created, and they are generally created to exploit new growth opportunities. But not every new endeavor necessitates a new firm. In fact, existing firms, as expected, exploit many new opportunities because they generally have access to a much better pool

¹³ Consistent with my claim, Telerate, a competitor interested in acquiring the expertise of Jeff Parker, made an offer to buy out TDC at a price that did not reflect the value of the growth options. Jeff Parker rejected the offer, developed the growth options, and sold out three years later at seven times as much.

of resources. Thus, new firms are created to exploit growth options that existing firms cannot or do not want to exploit. But what are these options? A theory of the ultimate causes of “governance overhang” in organizations should be able to illuminate us on this front as well.

VI. Conclusions

This paper makes an extremely simple point. *Corporate* finance is the study of the way *firms* are financed. Thus, our view of what a firm is shapes the way we do research in the field and the policy implications we derive. This influence has always been present but was less strong in the past because firms were very asset intensive and their boundaries were sufficiently stable that one could safely ignore the impact financing and governance choices had on them.

This is not a sensible approach anymore. In human capital-intensive firms, financing and governance choices can easily change the boundaries of the organization itself. As this type of firm becomes more and more common, the fundamental questions in theory of the firm become central to the corporate finance debate. How does an organization succeed in acquiring power that differs from ordinary market contracting? How is this power maintained, enhanced, or lost? How is this power used in a way that differs from ordinary market contracting? How is the surplus generated by the firm divided among its members? Answering these questions has become a precondition for any further advancement in corporate finance. I present an example of how such an approach can provide fruitful implications in the three main areas of corporate finance: capital structure, corporate governance, and valuation.

REFERENCES

- Aghion, Philippe, and Patrick Bolton, 1992, An incomplete contract approach to financial contracting, *Review of Economic Studies* 59, 473–494.
- Alchian, Armen, and Harold Demsetz, 1972, Production, information costs and economic organization, *American Economic Review* 62, 777–705.
- Allen, Franklin, and Douglas Gale, 1994, *Financial Innovation and Risk Sharing* (MIT Press, Cambridge, Mass.).
- Baker, George, Robert Gibbons, and Kevin J. Murphy, 1999, Relational contracts and the theory of the firm, Working paper, Harvard Business School.
- Bhattacharya, Sudipto, 1979, Imperfect information, dividend policy, and the “bird in the hand” fallacy, *Bell Journal of Economics* 10, 259–270.
- Bhide, Amar V., 2000, *The Origin and Evolution of New Businesses* (Oxford University Press, New York,).
- Black, Fisher, and Myron Scholes, 1973, The pricing of corporate liabilities, *Journal of Political Economy* 81, 637–659.
- Bolton, Patrick, and David Scharfstein, 1990, A theory of predation based on agency problems in financial contracting, *American Economic Review* 80, 93–106.
- Bulow, Jeremy, and John Shoven, 1978, The bankruptcy decision, *Bell Journal of Economics* 9, 437–456.
- Chandler, Alfred, 1977, *The Visible Hand* (Belknap Press, Cambridge, Mass.).
- Chandler, Alfred, 1990, *Scale and Scope* (Belknap Press, Cambridge, Mass.).

- Coase, Ronald, 1937, The nature of the firm, *Economica* 4, 386–405.
- Coase, Ronald, 1990, *The Firm, the Market, and the Law* (University of Chicago Press, Chicago).
- Diamond, Douglas, 1989, Reputation acquisition in debt markets, *Journal of Political Economy* 97, 828–861.
- Dybvig, Phillip, and Jaime Zender, 1991, Capital structure and dividend irrelevance with asymmetric information, *Review of Financial Studies* 4, 201–219.
- Economist*, 1998, A brand new strategy, November 21, 10.
- Everett v. Phillips et al.*, Court of Appeals of New York, June 4, 1942, 43 North Eastern Reporter, 2nd series.
- Fama, Eugene, 1990, Contract costs and financing decisions, *Journal of Business* 63, 71–92.
- Fama, Eugene, and Michael Jensen, 1983, Separation of ownership and control, *Journal of Law and Economics* 26, 301–325.
- Fama, Eugene, and Merton H. Miller, 1972, *The Theory of Finance* (Dryden Press, Hinsdale, Ill.).
- Fluck, Zsuzsanna, 1998, Optimal financial contracting: Debt versus outside equity, *Review of Financial Studies* 11, 383–418.
- Grossman, Sanford, and Oliver Hart, 1980, Takeover bids, the free rider problem, and the theory of the corporation, *Bell Journal of Economics* 11, 42–69.
- Grossman, Sanford, and Oliver Hart, 1986, The costs and the benefits of ownership: A theory of vertical and lateral integration, *Journal of Political Economy* 94, 691–719.
- Hamel, Gary, and C. K. Prahalad, 1994, *Competing for the Future* (Harvard Business School Press, Cambridge, Mass.).
- Hansmann, Henry, 1996, *The Ownership of Enterprise* (Belknap Press, Cambridge, Mass.).
- Harris, Milton, and Artur Raviv, 1991, The theory of capital structure, *Journal of Finance* 46, 297–355.
- Hart, Oliver, and John Moore, 1989, Default and renegotiation: A dynamic model of debt, Working paper, Department of Economics, MIT.
- Hart, Oliver, and John Moore, 1990, Property rights and the nature of the firm, *Journal of Political Economy* 98, 1119–1158.
- Hart, Oliver, and John Moore, 1994, Debt and the inalienability of human capital, *Quarterly Journal of Economics* 109, 841–879.
- Hellmann, Thomas, and Manju Puri, 1999, Venture capital and the professionalization of start-up firms: Empirical evidence, Working paper, Stanford University.
- Jensen, Michael C., and William Meckling, 1976, Theory of the firm: Managerial behavior, agency costs and capital structure, *Journal of Financial Economics* 3, 305–360.
- John, Kose, and Joseph Williams, 1985, Dividends, dilutions, and taxes: A signaling equilibrium, *Journal of Finance* 40, 1053–1070.
- Kaplan, Steven K., and Richard Ruback, 1995, The valuation of cash flow forecasts: An empirical analysis, *Journal of Finance* 50, 1059–1093.
- Kay, John, 1996, *The Business of Economics* (Oxford University Press, Oxford).
- Klein, Benjamin, 1988, Vertical integration as organizational ownership: The Fisher Body-General Motors relationship revisited, *Journal of Law, Economics and Organization* 31, 199–213.
- Klein, Benjamin, Alwin Crawford, and Armen Alchian, 1978, Vertical integration, appropriable rents and the competitive contracting process, *Journal of Law and Economics* 21, 297–326.
- La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer, 1999, Corporate ownership around the world, *Journal of Finance* 54, 471–518.
- Luherman, Timothy A., and William A. Teichner, 1992, Arundel partners: The sequel project, Case 9-292-140, Harvard Business School.
- Modigliani, Franco, and Merton H. Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *American Economic Review* 48, 261–297.
- Monks, Robert, and Nell Minnow, 1995, *Watching the Watchers: Corporate Governance for the Twenty-first Century* (Blackwell Publishing, Malden, Mass.).
- Myers, Stewart, 1977, Determinants of corporate borrowing, *Journal of Financial Economics* 5, 147–175.

- Myers, Stewart, 2000, Outside equity, *Journal of Finance*, forthcoming.
- Myers, Stewart, and Nicholas Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, *Journal of Financial Economics* 13, 187–221.
- Novaes, Walter, and Luigi Zingales, 1993, Financial distress as a collapse of incentives, Working paper, University of Chicago, Graduate School of Business.
- Novaes, Walter, and Luigi Zingales, 1998, Bureaucracy as a mechanism to generate information, Working paper, University of Chicago, Graduate School of Business.
- Parrino, Robert, and Michael Weisbach, 1999, Measuring investment distortions arising from stockholder-bondholder conflicts, *Journal of Financial Economics* 53, 3–42.
- Rajan, Raghuram, Henry Servaes, and Luigi Zingales, 2000, The cost of diversity: Diversification discount and inefficient investment, *Journal of Finance* 55, 35–80.
- Rajan, Raghuram, and Luigi Zingales, 1998, Power in a theory of the firm, *Quarterly Journal of Economics* 108, 387–432.
- Rajan, Raghuram, and Luigi Zingales, 1999, The firm as a dedicated hierarchy: A theory of the origins and growth of firms, Working paper, <http://gsblgz.uchicago.edu>, University of Chicago, Graduate School of Business.
- Rajan, Raghuram, and Luigi Zingales, 2000a, *The Governance of the New Enterprise in Corporate Governance*, X. Vives, ed. (Cambridge University Press), forthcoming.
- Rajan, Raghuram, and Luigi Zingales, 2000b, The tyranny of inequality: An inquiry into the adverse consequences of power struggles, *Journal of Public Economics*, forthcoming.
- Rose, Nancy, 1987, Labor rent sharing and regulation: Evidence from the trucking industry, *Journal of Political Economy* 95, 1146–1178.
- Ross, Steven, 1977, The determination of financial structure: The incentive signaling approach, *Bell Journal of Economics* 8, 23–40.
- Sahlman, William A., 1983, Technical Data Corporation, Case 9-283-072, Harvard Business School.
- Shleifer, Andrei, and Lawrence Summers, 1988, Breach of trust in hostile takeovers, in A. J. Auerbach, ed.: *Corporate Takeovers: Causes and Consequences* (University of Chicago Press, Chicago).
- Shleifer, Andrei, and Robert Vishny, 1992, Liquidation value and debt capacity: A market equilibrium approach, *Journal of Finance* 42, 1343–1365.
- Shleifer, Andrei, and Robert Vishny, 1997, A survey of corporate governance, *Journal of Finance* 52, 737–783.
- Stein, Jeremy, 1997, Internal capital market and the competition for corporate resources, *Journal of Finance* 52, 111–133.
- Titman, Sheridan, 1984, The effect of capital structure on a firm's liquidation decision, *Journal of Financial Economics* 13, 137–151.
- Welch, Ivo, 1997, Why is bank debt senior? A theory of priority among creditors, *Review of Financial Studies* 10, 1203–1236.
- Wernerfelt, Birger, 1984, A resource based view of the firm, *Strategic Management Journal* 5, 171–180.
- Williamson, O., 1985, *The Economic Institutions of Capitalism* (Free Press, New York).
- Zingales, Luigi, 1994, The value of the voting right: A study of the Milan stock exchange experience, *Review of Financial Studies* 7, 125–148.
- Zingales, Luigi, 1995, What determines the value of corporate votes? *Quarterly Journal of Economics* 110, 1047–1073.
- Zingales, Luigi, 1998a, Corporate governance, in Peter Newman ed.: *The New Palgrave Dictionary of Economics and the Law* (Stockton Press, London).
- Zingales, Luigi, 1998b, Why it's worth being in control, in George Bickerstaffe, ed.: *The Complete Finance Companion* (FT Pitman Publishing, London).