Toward an integrative theory of the firm: A relationship-based approach

Xin Li
Asia Research Centre
Copenhagen Business School
Tel.: (+45) 3815 3406
xl.int@cbs.dk

Abstract

This paper argues that the existing four major theories of the firm, i.e., the transaction cost theory, resource-based view, the entrepreneurial theory, and the stakeholder theory, are all insightful yet partial because each of them has a particular focus on the phenomenon of the firm. To better understand the nature and behaviors of the firm, we need a comprehensive yet integrative theory. Toward this end, this paper proposes a relationship-based theory of the firm (R’BT) which claims that it is the relationships between the entrepreneur and other individuals or firms that determine the existence, boundary, internal organization and competitive advantage of the firm.

Keywords: theory of the firm, transaction cost, resource, entrepreneur, stakeholder, relationship

1. Introduction

The theory of the firm is important not only to economists but also to business strategists as Porter (1994: 423) points out that ‘any effort to understand success must rest on an underlying theory of the firm and an associated theory of strategy’. In the extant literature, there exist four major theories, i.e., the transaction cost theory, the resource-based view (RBV), the entrepreneurial theory, and the stakeholder theory.

These four theories offer very distinct explanations of the nature of the firm and compete for supremacy. Each of them has a particular focus on phenomenon of the firm, making the field fragmented without an integrated theory. The transaction cost economics (TCE) focuses exclusively on reduction of transaction costs, including agency cost, costs of shirking, opportunism and underinvestment. The resource-based view (RBV), including core competence view (Prahalad and Hamel, 1990), knowledge-based theory (Kogut and Zander, 1992), and dynamic capability view (Teece, Pisano and Shuen, 1997), puts major emphasis on pursuing synergy derived from combined resources, knowledge and capabilities that create value. The entrepreneurial theory examines how entrepreneurs delegate judgmental decisions to others and process information within the firm. The stakeholder theory cares about the normative issue of wealth distribution, namely, whether the interests of all stakeholders involved in and affected by the operation of the firm are fairly satisfied.

Some scholars, as Foss (1999) points out, have taken an ‘isolationist’ approach in conducting research in theory of the firm, namely, they believe each tradition can develop a full-blown theory of the firm. I support Foss’s (1999) argument that an ‘integrationist’
approach is needed if we want to develop a comprehensive yet integrative theory. Since delegating entrepreneurial decisions, reducing costs, increasing value-added and distributing benefits generated by the firm are all indispensable aspects of the firm, an integrative theory of the firm must deal with all of them simultaneously. In this paper, I argue that the existing theories, though partial, offer some valuable insights which provide a basis for building an integrative theory of the firm. Building on existing theories, I propose a relationship-based theory of the firm to integrate existing theories.

The following text is organized as follows. In section two, I briefly review the four existing theories of the firm. Then I highlight some problems of the existing theories. In section four, I propose my relationship-based theory (R’BT) as an integration of the existing theories. Section five concludes the paper.

2. The four major theories of the firm

2.1 The transaction cost theory

Before 1930s, the neoclassical ‘theory of the firm’ was merely a part of the wider theory of value, which was constructed ‘for the purpose of assisting in the theoretical investigation of one of the central problems of economic analysis – the way in which prices and the allocation of resources among different uses are determined’ (Penrose, 1959: 11). Due to such a specific purpose, the firm in the neoclassical economics is treated as an atomic entity or a ‘representative firm’ that behaves merely in response to the price signals. Few economists thought it necessary to enquire what happened inside the firm (Penrose, 1995: x) and the firm either remained a black box or had no ‘insides’. In a sense, the neoclassical ‘theory of the firm’ is actually the theory of markets. Such a ‘black box’ treatment would ultimately invite criticisms from those who are concerned about the nature and behaviors of the real firms.

Coase (1937) was one of the first who dared to challenge the neoclassical theory of the firm and enquired into the nature of the firm. The question inspired Coase’s theorization is that, if the market/price mechanism were so efficient, why did the firm come to exist at all? Coase’s own answer is that, there are costs of using the market as well as the firm, i.e., cost of transacting and cost of organizing. The firm exists to reduce the transaction costs of searching and using market exchanges for everything it needs for production. However, there will also be costs of organizing the activities in the firm occurred whenever one extra transaction is internalized. The firm will tend to expand its size until the increase in organizing cost is equal to or greater than the reduction of transacting cost. Regarding the issue of internal organization of the firm, Coase argues under the employment relationship, the owner of the firm has the rights to control the works of the employees by telling them when to work, what to do and how to do it. Such an institutional authority is termed as fiat by Williamson (1971: 113-114). Williamson also argues the firm can act as a comparatively efficient resolution machinery to settle minor conflicts by fiat with authority.

In spite of the merits of Coase’s transaction cost theory, it has not made a substantial

---

1 Foss (1998: 9) points out ‘Coase had given virtually no details on the determinants of these costs’.
impact on the mainstream economic thinking before 1970s, which Coase lamented later that his theory was ‘much cited and little used’ (Coase, 1972: 68). There are several possible reasons. One might be Coase’s verbal theory could be hardly put into mathematical forms or in Williamson’s (1975) term ‘not been made operational’, so economists would not be willing or able to use it for the sake of publication, and Duncan Black’s (1958) was certainly the case (Coase, 1988: 36). Another reason might be his theory was overshadowed by the monopolistic competition revolution in economics led by Chamberlin (1933) and Robinson (1933) and later Debreu’s (1959) general equilibrium theory in his *Theory of Value*.

Entering into 1970s just about the time when Coase made his lamentation, there began to emerge some serious works on the theory of the firm. The seminal contributions which have defined the central streams of what is now called modern economic theories of the firm include Williamson’s (1971, 1975, 1985) transaction cost theory of vertical integration, Alchian and Demsetz’s (1972) team production theory, Marschak and Radner’s (1972) information processing theory of teams, Ross’s (1973) principal-agency theory, Nelson and Winter’s (1973) evolutionary theory of the firm, and Jensen and Meckling’ (1976) nexus of contracts theory. In late 1980s there emerged another important stream of theory of the firm – new property rights theory – developed by Grossman and Hart (1986) and Hart and Moore (1990). In a sense, most of these works except the information processing and evolutionary theories can be collectively termed as transaction cost economics (TCE) because they are all inspired by Coase’s transaction costs logic, and the key concepts in their analyses such as agency cost, cost of shirking, cost of opportunism, and cost of underinvestment are but different kinds of transaction cost.

2.2 The resource-based view

As Coase himself later recognized (Coase, 1988: 47) that while his theory ‘emphasized the comparison of the costs of transacting with the cost of organizing’, he ‘did not investigate the factors that would make the costs of organizing lower for some firm than for others’, i.e., the firm-specific factors which give one firm an competitive advantage over another firm. Evidently, this omission prevails in most of the economic researches on the firm. In response, many economists outside the mainstream and scholars of organization and management studies have tried to address the question what are the sources of firm heterogeneity or competitive advantages. In Conner’s (1991: 139) terms, we can treat the firm as not only an ‘avoider of a negative’ but also a ‘creator of a positive’.

In 1959, Edith Penrose (1959) published her landmark book *The Theory of the Growth of the Firm*. Although Penrose’s focus was on the growth of the firm and the limits on the growth rate, she nevertheless laid down the intellectual foundation of what was later called the resource-based view of the firm. Following Penrose, scholars who subscribe to the RBV treat the firm as a collection of productive resources. They have developed a variety of theories, such as, resource-based view (Barney, 1986, 1991; Wernerfelt, 1984), core competence view (Prahalad and Hamel, 1990), knowledge-based view (Grant, 1996; Kogut and Zander, 1992), dynamic capability view (Teece, Pisano, and Shuen, 1997),
among many others. Because most of these key concepts which are said to be sources of firm-specific competitive advantage can be seen as a sort of knowledge or capabilities the firm has for dealing with firm activities, some people often group them together under general umbrella labels such like knowledge-based perspectives or competence perspective (Williamson, 1999).

2.3 The entrepreneurial theory

Casson (2000, 2005) argues that existing theories of the firm are partial theories because they have largely neglected the entrepreneurial dimension of the firm, which is echoed by Foss and Klein (2005). According to Casson (2005), the neglect of entrepreneurship by the theorists of the firm may be due to two reasons. One is the traditional literature on entrepreneurship, such as Cantillon (1755) and Knight (1921), is economic and has little relevance to management of the firm. Another reason is that some of the economic literature on entrepreneurship involves distinctive psychological assumptions which though very insightful do not fit with mainstream theory. In addition, Foss and Klein (2005) find another reason, namely, many recent theories of entrepreneurship, such as entrepreneurship as imagination or creativity, entrepreneurship as innovation, entrepreneurship as alertness or discovery, and entrepreneurship as charismatic leadership, do not provide a natural links to the theory of the firm.

On the other hand, Witt (1999) points out that ‘theories focusing on the entrepreneurial element in the economy, in particular the contributions in the Austrian tradition, have paid hardly any attention to the role of the firm’. Therefore, there has been a missing chapter on the role of the firm in Austrian economics. Many people have made substantial efforts in bridging the theory of entrepreneurship and the theory of the firm, which result in an emerging entrepreneurial theory of the firm (see, Alvarez and Barney, 2004; Casson, 2000, 2005; Dew, Velamuri and Venkataraman, 2004; Foss, 1994, 1997; Foss and Klein, 2005; Foss, Foss, and Klein, 2007; Langlois, 1988; Silver, 1984; Witt, 1999, 2001). An insightful argument in these entrepreneurship perspectives is that it is the market failure that calls for the formation of a firm by an entrepreneur. For instance, Fose and Klein (2005) argue that the market may not be able to evaluate entrepreneurial services, and therefore, the entrepreneur cannot be an employee, but will instead start his own firm. So the existence of the firm can be understood as to reduce a specific category of transaction costs that close the market for entrepreneurial judgment; and changes in boundary is therefore viewed as response to a valuation discrepancy. Similarly, Zander (2007: 1142) argues that the formation of a new firm, or internalization of markets, is ‘the result of the inability of other market participants to either accept or understand the entrepreneur’s subjectively perceived means-ends framework, which to his or her mind spells out the preferred or ‘best way’ of implementing the entrepreneurial idea in the marketplace’.

In agreeing with Langlois (2005) that ‘entrepreneurship, properly understood, is a crucial but neglected element in explaining the nature and boundaries of the firm’, Casson (2005) claims that ‘the addition of entrepreneurship is not a minor refinement of the theory of the firm, but represents a radical change’.

2.4 The stakeholder theory

Of the four major theories, the stakeholder theory is not a ‘theory of the firm’ in the strict
sense, as it does not specifically address the three fundamental Coasian questions, namely, the existence, boundary and internal organization of the firm. The stakeholder theory was initially developed by Freeman (1984) as a strategic management theory which argues that, if business organizations are to be successful in current and future environment, then executives must take multiple stakeholder groups into consideration because some of them may prevent our accomplishments. This stakeholder approach is designed in direct contrast to the shareholder theory, which is best described by Milton Friedman (1977) that the only social responsibility of business is to increase its profits, or as more widely expressed, to maximize the shareholder value.

As a philosopher, R. Edward Freeman has had less problem and fear to talk about value, ethics, and distributive justice. This is the reason that the stakeholder theory of the firm, since its inception, has inherited such a normative element. It is worth noting that there is no a single stakeholder theory. Indeed, as Donaldson and Preston (1995) have categorized, there exist three broad types of stakeholder theories, namely, normative, descriptive and instrumental stakeholder theories. Different from normative theories, descriptive stakeholder theory is used to describe and sometimes to explain specific corporate characteristics and behaviors, while instrumental stakeholder theory is used to identify the connections, or lack of connections, between stakeholder management and the achievement of traditional corporate objectives such as profitability and growth (Donaldson and Preston, 1995: 70-71).

Although the stakeholder theory has been not bothered addressing the three Coasian questions, it does have potential for doing so. For instance, with regard to the existence of the firm, Evan and Freeman (1988) declare ‘the very purpose of the firm is, in our view, to serve as a vehicle for co-coordinating stakeholder interests’. Also, Freeman’s (1984) now classical book is entitled as Strategic Management: A stakeholder approach, which indicates its focus on the competitive advantage of the firm, the fourth element of a ‘strategic’ theory of the firm (Foss, 1999). Admittedly, as Schilling (2000: 230) rightfully points out that the stakeholder theory does have problem in defining the boundaries of the firm.

3 The problems of the existing theories
3.1 The transaction cost theory

Scholars, mostly non-professional economists, have made four criticisms on the transaction cost economics (TCE). First, the transaction cost logic is tautological; second, Coase’s original theorization includes imprecise analyses; third, most of the transaction cost researches have adopted a strategy of indirect test; and the fourth also the most serious one, TCE and Williamson’s (1985) opportunism-based version in particular is very ‘bad’ for practice.

3.1.1 Tautology in transaction cost logic

Masten, Meehan and Snyder (1991) have noticed that there is a tautological formulation built into the transaction cost arguments, namely, the central hypothesis underlying TCE is that the choice of which governance mode (i.e., market or the firm) to choose depends completely on the comparison between the costs of organizing under the corresponding
alternative. However, the direct comparison is impossible for two reasons. First, such costs are difficult to observe and measure given some implicit or latent costs exist; second, the cost of organizing under the alternative mode cannot be directly observed and measured because once one mode is chosen the alternative mode is simultaneously forgone. So, Masten et al. (1991) make it crystal clear that Coase’s original transaction cost argument is easy to make and impossible to refute. Barney (2001) echoes this charge and labels it as ‘Coasian tautology’. To borrow Williamson’s (1999: 1093) notion of ‘ex post rationalization’, we can argue the transaction theorists in the end can only say ‘show me a transaction carried out in the firm or in the market or under any type of governance mode, I can tell you it is because the reduction in transaction costs is greater than the increase in coordination cost by using that mode in question’. Yes, they may say so, but they can never show you how and why this is the case because they have no idea about and will never know how much transaction cost is reduced and how much coordination cost increased.

3.1.2 Imprecise explanations

Klein (1983: 373) criticizes Coase being mistaken to make a sharp distinction between intrafirm and interfirm transactions by claiming ‘economists now recognize that such a sharp distinction does not exist’. Klein suggests considering all transactions (within or between firms) as contractual (market) relationships. To Klein, what is the essential characteristic of a firm now appears to be unimportant, an idea held precisely by Richardson (1972) who sees a continuum ranges from pure market transaction through business network to pure firm-based planning. In a sense, both Richardson and Klein support the nexus-of-contract view of the firm of Jensen and Meckling (1976).

Zajac and Olsen (1993) argue that transaction cost perspective has at least two limitations when used to analyze interorganizational strategies: (1) a single-party, cost minimization emphasis that neglects the interdependence between exchange partners in the pursuit of joint value, and (2) an over-emphasis on the structural features of interorganizational exchange that neglects important process issues. They propose instead a transactional value framework for analyzing interorganizational strategies that emphasizes joint value maximization. Dyer (1997) echoes this view and contends that the fundamental governance question is ‘how can exchange relations be structured to maximize transaction value’ rather than ‘to economize on transaction costs’?

Also, I argue transaction cost theory cannot precisely explain the emergence of the firm because the reality is: when the entrepreneur decided to set up the firm, he could not imagine what kind of organizing costs will incur, needless to say he was to compare whether the reduction of transaction costs is greater than the increase of coordinating costs; he may have some ideas about the cost of salaries for employees, but he was actually not sure whether his firm could make a profit or not; he even had no idea whether he could fully pay the employees; nevertheless, he was risk-taking and optimistic, so he was not very bothered by these costs issue (and the comparison), and what in his mind might be: don’t worry, just do it.

3.1.3 Indirect tests in transaction cost researches
In order to sidestep the direct comparison problem, as Masten et al. (1991: 3-4) observe\(^2\), the TCE scholars have switched to ‘relate the incidence of transaction costs to observable characteristics of the transaction and then base predictions of organization form on those observed features’, a common strategy is ‘to estimate organizational form as a function of observable characteristics such as the degree of asset specificity and the level of uncertainty or complexity associated with the transaction’, which is precisely how Williamson’s theory is developed.

However, Masten et al. (1991: 4) rightly point out ‘a finding that higher asset specificity leads to a larger probability of integration could, in principle, obtain even if the hypothesis that asset specificity raises contracting costs were invalid [italic added]’. And indeed, the hypothesis is invalid as Dyer (1996, 1997: 536) has shown by the evidences of Japanese automaker-supplier keiretsu that, compared with US counterparts, the Japanese transactors incurred significantly lower transaction costs even though they had made greater asset-specific investments; and more interestingly, the automaker with the least specialized supplier group (GM) had the highest transaction costs while the automaker with the most specialized supplier group (Toyota) had the lowest transaction costs.

In addition, Ghoshal and Moran (1996: 23) argues Williamson’s logic that hierarchical control can reduce opportunistic behavior is not justifiable because the net effect of hierarchical control on opportunistic behavior will depend in part on the relative strengths of two distinct and mutually contradictory effect, i.e., formal control can in some situation constrain opportunistic behavior and in some other situation can invite opportunistic behavior due to lack of trust and negative feeling for the management, and the directionality of this net effect cannot be theoretically determined.

3.1.4 TCE being bad for practice

Among the critics from outside economics field, Pfeffer (1994), Ghoshal and Moran (1996), and Hodgson (2004) have made the most serious attacks on transaction cost economics, especially on Williamson’s opportunism-based theory. They argue TCE and opportunism-based theory in particular is established on false or at least imprecise assumptions about human nature, i.e., being opportunistic. According to psychology theories, people’s attitudes and behaviors are likely to be modified in response to their perceptions how they are treated by other people surrounding them. Negative perception derived from mistrust due to the opportunism assumption may reinforce people’s opportunistic attitude and behaviors through a positive feedback loop. Therefore, the opportunism becomes a self-fulfilling prophecy. Due to the self-fulfilling propensity of the opportunism-based TCE theory, it is very dangerous for management practice to design organizational control and incentive mechanisms according to this theory. The irony of the opportunism-based TCE theory is that it is likely to cause more opportunistic behaviors that the theory is designed to reduce.

Based on the above criticisms, I argue we should reject the transaction cost argument as a canonical theory of the nature of the firm for two reasons: firstly, the ambiguity in the

\(^2\) Foss (1998: 16) also notices ‘formal contributions prefer in general not to use the concept [transaction costs] at all (or only in a verbal introduction) and to model not transaction cost per se, but rather the manifestations of transaction costs’.
terminologies of transaction costs and coordination costs, i.e., lack of definitions and measurements of these costs and thus impossibility of comparison between them, is sufficient to ‘call into questions of the entire edifice’ (Ghoshal and Moran, 1996: 23) of transaction cost economics. Secondly, transaction cost theory as an economic theory, by focusing solely on costs, completely neglects other aspects of commercial exchange, i.e., the value side and the distribution of value created, which are two fundamental questions concerning the theory of the firm (Asher, Mahoney and Mahoney, 2005). In this sense, the resource-based view of the firm provides a better perspective on how value is created while stakeholder theory deals with the distribution of value created issue. Thirdly, Williamson’s opportunism-based theory of transaction cost is a negative theory and bad for practice.

3.2 The resource-based view

There are three major problems in the RBV of the firm, namely, the lack of robustness, being inward looking, and imprecise analysis.

3.2.1 The lack of robustness

Williamson’s (1999) critique on core competence theory as lacking of apparatus therefore the theory can only offer ‘ex post rationalization’. Namely, the core competence theorists tend to say, “show me a success story, I will identify a core competence; or show me a failure story, I will specify a missing competence”. We might be able to argue this ex post rationalization critique can be applied to the broader group of the resource-based view, especially with reference to Barney’s (1991) ‘valuable, rare, inimitable, and unsubstitutable’ (VRIN) framework. Williamson (1999: 1094) and Winter (1995: 149) charge the knowledge-based perspective have an expansion tendency, namely, its theorists tend to commoditize the concept of knowledge, capability, resource and routine, etc. to include so many different things, which in turn make it almost impossible to ‘breathe operational content into such competence features’ which Williamson (1999: 1097) sees a need for those competence perspective holders.

Foss (1998: 19) argues the information processing theories and knowledge-based (or competence-based) theories have a common problem: all incentive conflict problems are suppressed in order to focus on the costs of storing, using, producing and transmitting information and knowledge. This is very unrealistic given the undisputable existence of conflict of interests in our economies and societies. Also, Foss argues that the competence perspective (using Williamson’s (1999) term) has largely originated in the outskirts of economics such as business strategy, international business, technology studies, etc., and therefore is considerably looser than most mainstream economic theories of the firm. If they are best for explaining the competitive advantage of the firm, they are not robust to deal with the existence and boundary issues although most of the knowledge-based theorists often strive to address those issues.

3.2.2 RBV is inward looking

As the founder of the resource-based view of the firm, Penrose’s (1959: 5) emphasis in all of her discussion was on ‘the internal resources of a firm’ as she believes that a firm’s rate of growth is limited by the growth of knowledge within it (Penrose, 1995: xvii).

---

3 Ghoshal and Moran (1996: 23) mainly target Williamson’s version of TCE.
Therefore, since Penrose (1959), there has been an inward looking character in the resource-based tradition of inquiry into the firm, which is recognized by Lavie (2006). However, as many strategy scholars have realized that, the ownership of resources is though sufficient but not a necessary condition for competitive advantage, making the RBV a less ideal theory of competitive advantage for firms in alliances (Addas and Pinsonneault, 2007). Mathews (2002) argues that, although many latecomer firms from developing countries lack initial advantages/resources, they can nevertheless succeed in accelerated internationalization by establishing linkages with resource-rich foreign firms and leverage the linked firms’ resources.

Another problem of such inward looking is its neglect of other external factors influencing firm performance, which is well described by Collis and Montgomery (1995):

“With the appearance of the concepts of core competence and competing on capabilities, the pendulum swung dramatically in the other direction, moving from outside to inside the company [...] The external environment received little, if any, attention, and what we had learned about industries and competitive analysis seemed to disappear from our collective psyche” (pp. 59-60).

3.2.3 The firm does not own any resources

Lavie (2006) points out that the RBV is based on the assumption that firms are heterogeneously endowed with idiosyncratic resources that are owned or controlled within their boundaries. However, this ownership and control view is an imprecise analysis if one considers the ‘legal fiction’ view of the firm of Jensen and Meckling (1976), according to whom, most organizations are simply legal fictions which serve as a nexus for a set of contracting relationships among individuals. Here, the individuals involved in the nexus of contracts are the real owners or providers of resource of all kinds. For example, the shareholders are the owners or providers of financial resources; the employees are the owners and providers of their labors, be it manual or intellectual; the suppliers are the owners and providers of materials or services supplied; and so on.

In such a nexus of contracts, the firm as a ‘legal fiction’ does not actually own any resources. It does not own the employees although it has large degree of authority on how to use the labors of the employees. The firm also does not own any financial resources because all financial assets are simply equity belonging to the shareholders. The so-called retained profits are simply undistributed dividends to the shareholders who in turn have the ultimate ownership. If the firm does not own any resources, then it is the relationships between the firm and the resource holders that determine whether the latter are willing to sell the use rights of their resources in exchange for monetary returns. If the relationship is maintained well, then the exchange will continue, otherwise, the resource holders can cease providing their resources to the firm.

3.3 The entrepreneurial theory

I argue that the entrepreneurial theory has three problems. The first is that, albeit insightful, the idea of entrepreneurship as information processing or judgment falls short of an ideal theory of the firm for a simple reason, namely, these two entrepreneurial theories fail to elucidate why on earth the employees would be willing to cooperate with the entrepreneur in processing information and making judgment as these tasks are
certainly a kind of responsibility rather than recreation.

The second is that both the TCE and the RBV have an implicit assumption of entrepreneurship. In the TCE, it is an implicit assumption that it is the entrepreneur who will realize the problem of transaction costs and desire to avoid such transaction costs. In the RBV, Penrose (1959) actually explicitly discusses the role of entrepreneurial services in the growth of the firm although she does not explain the link between the entrepreneur and the emergence of the firm. Therefore, although existing theories of the firm often do not talk much about the role of entrepreneur in the firm, the explication of such an implicit assumption per se does not give the newly proposed entrepreneurial theory enough legitimacy for claiming as a new alternative theory of the firm.

The third problem is that the delegation of entrepreneurial judgments cannot be the defining feature that distinguishes the firm-based relationships from the market-based relationships. It is established in the entrepreneurship literature that the defining character of an entrepreneur is that his or her unwillingness to work for others. This means, compared to non-entrepreneurs who will be willing to be employed by others, entrepreneurs will tend to work on their own as sole proprietors or employ others to work for them. If we treat sole proprietorship as a single-person firm, which is so in legal terms, then we can say that entrepreneurship existed in pure market transactions before the emergence of the firm, i.e., business organizations that employ no less than one employee. In such a time, a sole proprietor, who specializes in his or her own trade, actually ‘delegates’ other entrepreneurial judgments to other sole proprietors to decide how to produce the things the former wants. Without the delegation of entrepreneurial judgments, all the trading sole proprietors will have to live in a self-sufficient autarky in which every individual produces whatever he or she wants to have. In this sense, we can say that the delegation of entrepreneurial judgment is the determinant that makes the transition from self-sufficient autarky economy to market-based economy. However, it is not the determinant that triggers the change from a sole proprietorship to a modern firm employing no less than one employee.

3.4 The stakeholder theory

There are two reasons why the stakeholder theory alone cannot be a complete theory of the firm. First, the stakeholder theory was established initially as a strategic management theory with a strong normative standing, i.e., being not aimed as a theory of the firm. So, it does not bother explaining why the firm comes to exist, the fundamental question regarding the nature of the firm according to Coase (1937). Second, the stakeholder theory has so far downplayed the role of entrepreneur and his or her self-interests, which makes the stakeholder theory unrealistic to many people who are familiar with the conventional wisdom of shareholder theory.

However, I posit that the stakeholder theory, which sees the firm as a web of firm-stakeholder relationships [italic added] (Freeman, 1984), can be transformed into an alternative theory of the firm if integrated with the other insightful elements of the other three theories, i.e., the TCE, RBV and the entrepreneurial theory.
4 The relationship-based theory of the firm: an integration
4.1 The integration of existing theories

In this section, I propose a relationship-based theory of the firm (R’BT), which I argue can integrate the existing four theories (see Figure 1 for a visual representation). The R’BT integrates the entrepreneurial theory because it subscribes to the view that every firm is established by an entrepreneur who discovers or intentionally creates an entrepreneurial opportunity for wealth generation or realization of a novel idea. The entrepreneur is aware that his/her own resources (money, labor and knowledge) alone are not adequate to exploit such an entrepreneurial opportunity, so he/she needs to use others’ resources. By using the combined resources of the entrepreneur’s own and others, wealth will be generated, which in turn results in the issue of wealth distribution. How the wealth generated by a group of people will be distributed depends on the relationship orientation of the entrepreneur. If the entrepreneur is self-centric, then he/she may tend to see others opportunistic; so, to maximize his/her own self-interests, the entrepreneur tends to internalize the joint wealth creation activities (i.e., market exchanges involved in the process of wealth creation). In sharp contrast, if the entrepreneur is allo-centric, i.e., having group consciousness, he/she may tend to consider the interests of others and see the collaborations with them as opportunities for synergy. No matter choosing internalization or collaboration, the entrepreneur will learn continuously, which may help him/her to discover or intentionally create new entrepreneurial opportunities for next round wealth generation and distribution.

Figure 1: An integrative relationship-based model of the firm

Of course, there are other situations that are more complex than the simplistic dichotomy of self-centrism and allo-centrism can capture. In fact, they are just the two extremes of a continuum. If we import the means vs. the end terminology, we can identify at least four types of people who manage relationships differently (see Figure 2). First, if a person is self-centric in his/her means as well as ends, then he/she is opportunist (in Williamson’s (1985) term). Second, if a person is allo-centric in both means and ends, then he/she is altruist. Third, if his/her ends are self-centric while his/her means are allo-centric, then this person may be called strategist. Fourth, in contrast to the strategist, the pragmatist is allo-centric in his/her ends while self-centric in his/her means. It is worth noting that a person may be opportunist, strategist, pragmatist, or altruist in different circumstances (time, place and issues).
One may ask why the other individuals involved in the joint wealth creation via market exchanges are willing to be internalized by the entrepreneur. This is in fact a crucial question in a theory of the firm, and especially in the theory of the existence of the firm.

The TCE offers a partial answer that the entrepreneur wants to avoid the transaction costs associated with activities such as negotiation and monitoring. It is partial because it does not explain why the others would be willing to be employed if the only purpose for internalization/employment is to help the entrepreneur avoid or minimize his transaction costs. What the TCE neglects is the other party of the employment relationship, i.e., the employed. There must be clear benefits that help the employee-to-be to accept the proposed change from a market-based exchange relationship to a firm-based employment relationship. In essence, for the internalization/employment to happen, the two parties involved must perceive both of them will be better off with such a change of relationship. Therefore, it is the perceived mutual benefits derived from such a relationship change, rather than the avoidance of transaction costs on the part of the entrepreneur, that result in the real actions of the internalization/employment.

The RBV also offers a partial answer that the entrepreneur wants to pool together the resources of other people involved to generate synergy. It is a partial explanation because the synergy will be generated as long as these individuals work ‘together’ even if through market exchange. The knowledge-based theory (Kogut and Zander, 1992) argues that some kinds of knowledge creation demand close interactions and the firm provides a platform for such close interactions. However, this synergy pursuing argument still falls short in explaining why co-locations of knowledge generating individuals is not sufficient or preferable for the synergy purpose. In practice, there are examples that a firm has established long term contractual relationships with some independent contractors who work almost solely for the firm as if they were normal employees. Again, I argue, it is the perceived mutual benefits derived from such a relationship change, rather than the pursuit of synergy on the part of the entrepreneur, that result in the real actions of the internalization.

Then, what are the answers of the R’BT to the four fundamental questions regarding the nature of the firm? I will address them in the following sections.
4.2 The existence of the firm

Zander (2007) argues that existing theories except few have been mainly concerned with boundary decisions in a context of established firms and therefore largely neglected the question of new firm formation. I want to go one step further to enquire into the formation of the first ever firm in human history. Why and how did the first ever firm that employed no less than one employee (excluding the sole proprietorship that owned and run by the entrepreneur-owner) come to exist? To do so, we can analyze the evolution of human society by looking at the dominant mode of economic production in the society. As shown by Figure 3, the human society had been in self-sufficient autarky state before it entered the age of market economy, i.e., specialization and trading. The age of market economy can be further divided into two phases. The first phase was a sole proprietorship-based economy while the second (also the current) phase is the firm-based economy. Here, I refer the ‘firm’ as a business entity that employs no less than one employee, meaning that I don’t count sole proprietorship as a normal firm in the sense of employing organization.

As I have analyzed in section 3.3, entrepreneurship, i.e., specialization and delegating entrepreneurial judgmental decisions to other individuals, is the factor that made human society to shift from self-sufficiency to market economy. Then, what triggered the shift of human society from the sole proprietorship-based phase to the firm-based phase of market economy? My own answer is as follows. In some point in time, there was an entrepreneur/sole proprietor started to want to internalize the established market-based exchange relationships due to his desire of either avoiding transaction costs or pursuing synergy in resources held by different sole proprietors or both. With his/her internalization idea, he/she approached to one or more people (sole proprietors) who have established exchange relationships with the former. After persuasions and negotiations, both of them realized there would be mutual benefits to gain if they change their previous market-exchange relationships into firm-employment relationships. Once all of them who accepted such a relationship change took actions of internalization, the firm came to exist.

The entrepreneur or the firm may continue to internalize more and more previously market-based exchange relationships into firm-based employment relationships. The most extreme case is that all market-based transactions were to be internalized into one big firm. If that happens, then the whole economy will cease to be a market economy but morph into a centrally planned economy, or communism. However, the market force, i.e., competition, merger and acquisition, is highly unlikely to bring forth the communism. This means, without strong political force to substitute market by plans, it seems no any single firm can have such competitiveness to kill all the other rivals.
4.3 The boundary of the firm

After the firm comes to exist, there is a tendency for expansion, i.e., further internalization of market. Besides internalizing some of the established market relationships (be with individuals or other firms), the firm may seek direct employment relationships with new people. As long as an employment relationship will bring perceived mutual benefits to the two parties, i.e., the firm (or the entrepreneur-owner) and the employee-to-be, such an employment relationship may start and continue; otherwise, either party may reject or terminate such a relationship, which means, the firm may fire the employee or the employee may resign and leave the firm.

Of course, whether one party of the relationship will decide to terminate its employment relationship with the other party will depend on its perception of how bad or good the relationship is. To borrow Simon’s (1951) terminology, if the perceived quality of the relationship is within the ‘zone of acceptance’ of one party, it will not go directly to terminate the relationship even though it might have felt the deterioration of the relationship. Only when the perceived quality of the relationship decreases and falls outside the ‘zone of acceptance’, one will terminate the relationship immediately.

In other words, the boundary of the firm is determined by the qualities of relationships involved. For an established market-based relationship, an entrepreneur/firm may be able to internalize it for the desire of either avoiding transaction costs or pursuing resource synergy. Therefore, the firm tends to grow in size and scope. The limit to the growth of the firm and the rate of growth seems to be the capacity of the entrepreneur (or the firm) to manage more and more internalized relationships to maintain acceptable qualities in those relationships. The firm will not grow indefinitely for two reasons. For one, the managerial capacities of managing relationships are limited and tend to grow more slowly than the growth of the number of internalized relationships (cf. Penrose, 1959). For another, with more and more relationships internalized, there will be multiple interests groups within the firm, and it is not an easy task to coordinate the multiple interests of these groups. One policy or decision may cause conflict of interests. While some people feel satisfied, others may perceive decreasing qualities of their relationships.
with the firm. Therefore, while the firm continues to internalize more and more market relationships, there will be some people quit the firm, reducing the speed of firm expansion or even reversing the expansionary tendency.

4.4 The internal organization

When talking about the internal organization of the firm, Coase (1937) actually only refers to the authority of the firm/entrepreneur over the employees. Here, I will also limit my analysis to the organizational authority issue. The R’BT offers an alternative viewpoint on authority in the firm. Rather than lean towards either Coasian fiat-type authority of the employer or Alchian and Demsetz’s (1972) view of no authority at all, the R’BT maintains that the owners or their representative managers do have authority over their employees or subordinates. But such authority has its limit. According to Simon (1951, 1995), the employer’s authority over its employees is restricted to employees’ ‘zone of acceptance’ under the assumption of utility maximization rather than the neoclassical assumption of profit maximization (Simon, 1995: 279) where ‘profit [consideration] is only one among a number of [considerations] and enters into most subsystems [of the firm] only in indirect ways’ (Simon, 1964: 21). And the real authority lies in the capacity for the employer, under the precondition of acceptable-quality relationship, to prioritize its organizational goals over its employees’ individual goals.

This prioritization capacity based authority indicates that if the employer abuses its bargaining power or if its fiat falls outside the employees’ ‘zones of acceptance’, the employees may choose to leave the firm, and then there will be no any authority over the employees-to-leave any more. But, if the employer firstly build and harmonize its relationships with its employees, then the employer can more easily prioritize its organizational goal over employees’ and the employees tend to accept the prioritization and authority.

The R’BT also does not prefer the formal control argued for by TCE scholars, especially by Williamson. In China where relationships and harmony are highly appreciated and cultivated, the management philosophy goes that it is not impossible to achieve a perfect type of control: governing without intervening, which can be understood as the principle of subsidiarity, namely, don’t interfere your subordinates (people), if you selected them, trust them, empower them, and encourage their creativity; and then they might get tasks done effectively, efficiently and creatively. If leaders at every level in the firm follow this ‘governing without interference’ principle to manage their subordinates who manage their own tasks, then the whole company can be governed without interference.

4.5 The competitive advantage

The TCE explains the competitive advantage of the firm in terms of its capabilities to reduce or economize on transaction costs (Williamson, 1991a, 1991b) while the RBV sees the source of competitive advantage as the value-creating internal resources the firm owns and controls. Both views are insightful yet partial as competitive advantage is often manifested itself by a superior value/cost ratio, which means both cost and value aspects need to be considered when it comes to competitive advantage. The R’BT subscribes to the argument of the stakeholder theory that good management of relationships with stakeholders is the ultimate source of the firm’s competitive edge because good relationships may result in both low transaction costs and high value creation (Dyer, 1997;
Li, 1998; Dyer and Chu, 2003).

Here, it is worth explaining the difference between the R’BT and the RBV on the source of competitive advantage. According to Penrose (1959: 78), a firm controls some resources, including physical, human, managerial and entrepreneurial resources and ‘the services that the resources will yield depend on the capacities [italic added] of the men using them’. I see two types of capacities here, one being the relational capacity while the other the technical capacity. The relational capacity refers to the capacity of the entrepreneur or management to maintain a good relationship with the resources under the firm’s control so that the resources will be willing to work productively for the firm. The technical capacity refers to the capacity of the entrepreneur and management to utilize the resources at their full productivities. Between these two capacities, the relational capacity is more important than the technical capacity for the following reason. If the entrepreneur or management has high relational capacities but low technical capacities, the entrepreneur or management can still rely on the employees under the firm’s control to self govern their own works; in other words, decentralization under good firm-employee relationships can still work well. However, if the entrepreneur or management has high technical capacities but low relational capacities, then while technically capable entrepreneur or management wishes to centralize decision making, the employees may not be willing to work wholeheartedly for the firm, and the result will be low overall performance.

A particular example is the ‘trustworthy reputation’ as a valuable resource. According to Dyer and Chu (2003: 57), trustworthiness lowers transaction costs and may be an important source of competitive advantage. However, the trustworthy reputation is first of all a result of a prior relationship building efforts. In addition, without continuous relationship building, established or inherited trustworthy reputation may be damaged or eventually lost. In short, good relationships come before good resources for bringing good performance.

To maintain good relationships with stakeholders, the firm has to implement a fair mechanism for wealth distribution among all stakeholders involved. Only when all stakeholders/parties who have contributed to the wealth generation are fairly rewarded, they will feel satisfactory and the relationships in question will be in good quality. The stakeholder satisfaction is very important for sustainable wealth generation. Take employee for example, only when the employees are satisfactory about their employer, will they work committedly, productively and creatively as reciprocity. In a sense, the employees’ commitment, productivity and creativity are the root source of the competitive advantage of the firm.

5 conclusion

In this paper, I have reviewed the existing four major theories of the firm, i.e., the transaction cost theory, resource-based view, the entrepreneurial theory, and the stakeholder theory, and highlighted the problems of each of them. I acknowledge that all of these four theories are insightful yet partial because each of them has a particular focus on the phenomenon of the firm. To make better sense of the extant literature and understand the nature and behaviors of the firm, I have proposed a relationship-based
theory of the firm (R’BT) as a way to integrate the existing theories. The R’BT claims that it is the relationships between the entrepreneur/firm and other individuals or firms that determine the existence, boundary, internal organization and competitive advantage of the firm.

Acknowledgement

This paper is a result of an ongoing research dating back to 2006 when I was a second year PhD in strategic management student at Royal Holloway college of University of London (RHUL). Although I quit without finishing my PhD study from RHUL in September 2007, the College’s financial support for my study through a fee-waiver scholarship is acknowledged. I carried on this research in my spare time during my PhD study at Copenhagen Business School (Feb. 2008—Jan. 2011). The generous financial supports for my study from the School and the Danish EAC Foundation are highly appreciated. I want to thank Hui Tan, Michael Mol, Peter Ping Li, Verner Worm, Nicholai Foss, and Jon Bingen Sande for helpful comments. All errors are my own.

Reference

*New York Times Magazine*, September 13, 1970

Theory’, *Academy of Management Review*, 21(1): 13-47


Hodgson, G.M. (2004) ‘Opportunism is not the only reason why firms exist - why an 
explanatory emphasis on opportunism may mislead management strategy’, 
*Industrial and Corporate Change*, 13(2): 401–418

agency cost and ownership structure’, *Journal of Financial Economics*, 3: 305-360

Klein, B. (1983) ‘Contracting costs and residual claims: The separation of ownership and 


Kogut, B. and Zander, U. (1992) ‘Knowledge of the firm, combinative capabilities, and 


entrepreneurial firm’, Working Paper 2005-27R, Department of Economics, 
University of Connecticut, USA.


Yale University Press


University Press: New York


Pfeffer, J. (1994) *Competitive advantage through people: Unleashing the power of the 

School Press, pp. 423-462