The nature of the firm: A relationship-based theory

Xin Li
Asia Research Centre
Copenhagen Business School
xl.int@cbs.dk
Tele: +45 3815 3406
Fax: +45 3815 2500

Abstract

This paper is intended to propose a relationship-based theory of the firm (R’BT), an alternative to the transaction cost theory. The R’BT takes relationship disharmony (rather than transaction costs) as its basic unit of analysis of the nature of the firm, and argues relationship disharmony derives from conflict of interests, a problem faced by any human organizations; to reduce the conflict of interests, the best way is to harmonize relationships by mutually accommodating or internalizing the other party’s interests into its own consideration; the degree of harmony or disharmony in the relationships will in turn determine the performance of the organization in question. This paper provides a completely different explanation from the transaction cost economics when addressing issues such like the existence, boundary and internal organization of the firm. Also discussed are its implications for organizational management and economy.

Keywords: theory of the firm, transaction cost, relationship, harmony, Confucian

1. Transaction cost economics: A brief review

1.1 Coase’s (1937) seminal work ‘The Nature of the Firm’

Before 1930s, the orthodox economics treated the firm as a single actor which competes in the market in response to the signals of market prices of the input factors and output products. The firm in economics textbooks was essentially a ‘black box’ that transforms inputs into outputs according to a production function or a set of productive possibilities. Therefore, the theory of the firm was better to be understood as a theory of market (Jensen and Meckling, 1976: 306) or being completely subordinated to price theory (Foss, 1998: 4).

Although the conventional production-function theory of the firm provided useful insights for understanding pricing and output decisions and how these vary with competitive conditions (Klein, 2000: 463), it was not useful at all to address the questions regarding what determines the existence, boundary and internal organization of the firm. Being bothered by the discrepancy between the apparent importance of the firm, namely ‘most resources in a modern economic system are employed within the firm’, and the ‘black box’ approach to the firm which lacks any substance but a mere name (Coase, 1991: 714), a bright and young British student Ronald Coase, in his third year of undergraduate study at London School of Economics, decided to open up the ‘black box’ and find out the answers to the above mentioned three questions. Then there came his well-known 1937 paper ‘The Nature of the Firm’ published on Economica.
In this path-breaking paper, Coase raises two puzzles: the first is ‘why was [the firm] needed if the pricing system [i.e., the market] provided all the coordination necessary?’ (cf. Coase, 1991: 715), and the second ‘why, if by organizing one can eliminate certain costs and in fact reduce the cost of production, are there any market transaction at all?’ (Coase, 1937: 394). Coase’s own answer is: 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 huge transaction costs\(^1\) of searching and using market exchanges for everything you need for a 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 internal organization issue, Coase argues under the employer-employee relationship, the owner of the firm has the rights, or institutional authority in Williamson’s (1971:113) term, to control the works of the employees by telling them when to work, what to do and how to do it, which is termed as fiat by Williamson (1971: 114) who also argued the firm can be a comparative efficient resolution machinery to settle minor conflicts by fiat with authority.

1.2 The development of transaction cost economics

In spite of the merits of Coase’s transaction cost theory of the firm, it has not made a substantial impact on the mainstream economic thinking before 1970s, which Coase lamented later his theory being ‘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) 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.

If Coase (1937) was mainly aimed at explain why firm exists (Coase, 1988) and what determines the boundary of the firm in passing, the TCE is intended to operationalize Coase’s verbal theory of transaction cost – what are the major causes of increase in transaction costs. While Alchian and Demsetz (1972) offer a slight different answer to the existence issue of the firm, others like Williamson, Grossman, Hart and Moore try to shed light on why vertical integrations occur and

\(^1\) Foss (1998: 9) points out ‘Coase had given virtually no details on the determinants of these costs’.
how integration should be done.

1.3 Nexus-of-contracts theory of the firm

Alchian and Demsetz (1972) criticizes Coase’s (1937) argument that the firm has fiat-type authority over its employees. They argue there is no difference between firing your secretary and firing your grocery because the employer-employee relationship and the customer-grocery relationship is no difference from an economic perspective: the two relationships are both subject to ongoing renegotiations. Jensen and Meckling (1976) support Alchian and Demsetz’s (1972) argument and articulate the nexus-of-contracts theory: the firm is actually a legal fiction characterized as a nexus of contracts. The difference between Alchian and Demsetz (1972) and Jensen and Meckling (1976) is the former see the team production problems are central to the firm (rather than Coase’s transaction costs argument?) therefore the firm exists to reduce such costs of shirking in a production team while the latter more focus on agency problems which exist not only within the firm but also with suppliers, customers, creditors, etc. Jensen and Meckling’s (1976) thesis, rather than a theory of the firm, is actually a theory of ownership structure which examines ‘the nature of the agency costs generated by the existence of debt and outside equity and demonstrate who bears these costs and why’ (p. 305). The nexus-of-contract approach to the firm problem is very influential and inspired many researches including Holmstrom (1982) which is very formal, i.e., mathematical. Another two articles published in the 1990s (Barzel, 1997; Jensen and Meckling, 1992) demonstrate the nexus-of-contracts view is far from dead (Foss, 1998: 26).

Klein (1983: 373) argues ‘it is misleading, however, to think that the major contracts that define an organization consist solely of the residual claim and decision allocation structure’. Foss (1998: 24) has made several critiques of this team-production theory, namely, first, it is not necessarily the monitor must be the employer of the firm (cf. Holmstrom and Tirole, 1989); second, it is not impossible that the employees can monitor each other; third, it not meaningless to speak of authority of the employer over employee; and the last, we seem to observe more firms in the real world that can be explained by team production, such as conglomerates. I argue, Alchian and Demsetz’s (1972) team production theory cannot explain why there exists a team for joint production in the first place. In response to this critique, they may trace back to the far early beginning when the first ever team in human history was formed voluntarily in which there was no planner who decided to have the team beforehand. Maybe this is the best defense for why there is a team without a leader in the first place. However, anthropologists might be able to help us to demonstrate that there was no such kind of arrangement for one single person to become the monitor of the team and therefore be granted the rights of residual claim.

1.4 Williamson’s asset specificity theory

Oliver E. Williamson is the central figure for operationalizing Coase’s transaction cost theory and the development of TCE. According to Williamson (1975, 1985), human nature is assumed to be characterized as bounded rationality, capacity for foresight, and opportunism, i.e., self-interest seeking with guile, and significant transaction costs accrue due to three key attributes of complex business contracts: frequency, uncertainty, and asset specificity, of which the last one has the most explanatory power in Williamson’s theory. It is the combination of such human nature and such contract attributes that leads Williamson to argue that specific assets open the door to opportunism (Foss, 1998: 35) which causes transaction costs and economizing on transaction cost is the ‘main case’ in explaining the nature of the firm – the firm is an alternative governance structure to the
market which can, under certain circumstances, economize on bounded rationality and mitigate the hazards that accrue to opportunism’ (Williamson, 1999: 1090).

Williamson disagrees with Alchian and Demsetz’s (1972) assertion that there is no difference between the two types of contracts: one being employment contract and another being customer-supplier contract. Williamson treats the firm acting as an internal court to settle within-firm disputes when these disputes are more firm-specific and the litigation in law court may not be the most efficient means. Williamson’s most widely recognized contribution to operationalizing Coase’s (1937) transaction cost theory has been his asset specificity theory. According to this theory, asset specificity will cause higher transaction costs due to opportunism such as post-contract hold-up problems, so in order to economize on such transaction costs the ultimate solution is vertical integration because buy bringing the supplier who faces such asset specificity problems into the firm, the firm can request it to invest in such relationship-specific investment without holding-up risks.

Critics have pointed out Williamson’s asset specificity theory adopted an indirect test strategy (Masten, Meehan, and Snyder, 1991: 3) based on an assumption that asset specificity has a positive correlation with transaction costs. Williamson’s theory is then an asset specificity theory of vertical integration rather than a theory of the firm which has nothing to do with transaction cost, if considering Masten et al.’s (1991: 4) critique ‘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’.

1.5 The new property rights theory

Although asset specificity is adopted as an important concept in their analyses, Grossman and Hart (1986) and Hart and Moore (1990) (thereafter GHM) have independently developed a new approach to vertical integration issue – the new property rights theory, different from the old property right theory (e.g., Alchian and Demsetz, 1973; Coase, 1960; Demsetz, 1967). Same as Williamson’s theory, GHM’s new property rights theory is mainly for understanding the integration decision. Say, in a business relationship, the two parties may identify a relationship-specific investment by their own firms, since this will be an investment with asset specificity from each side, if they are kept as independent firms, each side may under-invest in such relationship-specific asset, but if one party acquires another party, the acquiring party will fully invest while the acquired party might still underinvest. So, the efficiency will be enhanced if it is the party whose investment in asset specificity is more important than the other party for the joint benefits to acquire the other party. Regarding the boundary issue, the new property rights theory argues it is the rights to residual claim, i.e., ownership of assets, that determine the boundaries of the firm: a firm is defined as a collection of jointly owned assets (Foss, 1998: 37).

My critique of GHM’s property rights theory is that they neglected an important issue: whether the party whose relationship-specific investment is more important to the relationship than the other party’s is possible to buy up the other party. Intuition tells us often in a business relationship it is the smaller party whose relationship-specific investment is more important to itself and to the business relationship because for a small company a certain amount capital investment may be a big and very important investment decision while such an investment is just trivial for a huge firm who happen to have a business relationship with the small firm. According to GHM’s theory, it is better and more likely for the small firm to acquire the huge firm. This is unthinkable in normal situation!
2. Criticisms and alternative theories to TCE

2.1 The criticisms of transaction cost economics

Most of the criticisms listed here are from scholars outside economics. There are four serious criticisms placed on TCE: first, the transaction cost logic is tautological; second, most of the transaction cost researches have adopted a strategy of indirect test; third, Coase’s original theorization includes imprecise analyses; and the final and the most serious criticism being that TCE and Williamson’s opportunism-based version is very bad for practice.

Tautology in transaction cost logic

Masten et al.’s (1991) have noticed that there is a tautological formulation built into the transaction cost arguments: the central hypothesis underlying TCE is that the choice of which governance mode to choose (i.e., market or the firm) completely depends on the comparison between the costs of organizing under the corresponding alternative. However, the direct comparison is impossible for two reasons, one being such costs are difficult to observe and measure given some implicit or latent costs exist; another one being the cost of organizing cannot be directly observed for the alternative not chosen. 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 label 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 will do 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 is increased.

Indirect tests in transaction cost researches

In order to sidestep the direct comparison problem, as Masten et al. (1991: 3-4) observe², 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]’. 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.

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’.
Imprecise explanations

Klein (1983: 373) criticizes Coase being mistaking to make a sharp distinction between intrafirm and interfirm transactions by claiming ‘economists now recognize that such a sharp distinction does not exist’. He suggests to consider all transactions (within or between firms) as contractual (market) relationships. In his mind, what is the essential characteristic of a firm now appears to be unimportant. He supports the nexus-of-contract view of the firm. Klein (1983: 373) also criticizes the residual claim (lixin: A&D, 1972) and decision allocation structure (lixin: 1973 Team theory) views of the firm as misleading.

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 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 kinds of issues (including the comparison), and what in his mind might be: DON’T WORRY, JUST DO IT.

TCE being bad for practice

Among TCE’s critics not from economics field, Ghoshal and Moran (1996) and Pfeffer (1994) 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 psychological theories, people’s attitudes and behaviours 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 organizational management practice to design organizational control and incentive mechanisms according to this theory. The irony of the opportunism-based TCE theory is likely to cause more opportunistic behavior they 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 altogether for two reasons: firstly, the ambiguity in the 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 Moran3, 1996: 23) of transaction cost economics.

2.2 Knowledge-based perspectives

As Coase himself recognized (Coase, 1988: 47) that while his theory ‘emphasized the comparision 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

---

3 Ghoshal and Moran (1996: 23) mainly target Williamson’s version of TCE.
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. These scholars have developed a very wide scope of theories, such as evolutionary economics (Nelson and Winter, 1973, 1982), resource-based view (Barney, 1986, 1991), core competence view (Prahalad and Hamel, 1990), knowledge-based view (Grant, 1996), dynamic capability view (Teece, P 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 the firm has in dealing with firm activities, so some people often group them together under general umbrella labels such like knowledge-based perspectives, resource-based view, or competence perspective (in Williamson’s (1991) term).

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.

Williamson’s (1999) critique on core competence theory as lacking of apparatus therefore the theory can only offer ‘ex post rationalization’: show me a success story, I will show you (uncover) a core competence; or show me a failure story, I will show you (uncover) a missing competence. We might be able to argue this ex post rationalization critique can be applied to the broader group of the knowledge-based perspective. Williamson (1999: 1094) and Winter (1995: 149) charge the knowledge-based perspective have a 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.

2.3 Entrepreneurial theories of the firm

Casson (2000, 2005) argues 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 entrepreneurship literature on entrepreneurship, such as Cantillon (1755) and Knight (1921), is economic and has little relevance to management of the firm. Another reason is some of the economic literature on entrepreneurship involves distinctive psychological assumptions which while very insightful do not fit with mainstream theory. In addition, Foss and Klein (2005) give another reason for such neglect as 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.

Casson, Foss and Klein, among others have made significant progresses in bridging the theory of
entrepreneurship and the theory of the firm by redefining entrepreneurship as information processing or managing (Casson, 2000, 2005) and entrepreneurship as judgment (Foss and Klein, 2005; Foss, Foss, and Klein, 2007). For instance, the entrepreneurship as judgment theory argues, the market may not be able to evaluate entrepreneurial services, 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.

Albeit insightful, I still consider these entrepreneurship as information processing and judgment approaches fall short of an ideal theory of the firm for two simple reasons: firstly, these two entrepreneurial theories of the firm fail to elucidate why on earth the employees would be willing to cooperate with the entrepreneur in processing information and making judgment as these are certainly a kind of responsibility rather than recreation; secondly, these theories fail to shed light on a crucial yet neglected question: why do people voluntarily come together to form the firm as co-founders and why do other people come to join the firm as employees in the first place?

3. The need for a new theory of the firm

3.1 Coase’s (1988) call for a comprehensive theory

Coase (1988:37) admits he was aware that there was a weakness in his exposition in his 1937 paper at the time the article was published, but it took him another 50 years to realize how serious that weakness may have been for the development of the subject, namely, in ‘The Nature of the Firm’, Coase presented ‘an incomplete picture of the nature of the firm’ (p. 37) by ‘concentration on the firm as a purchaser of the inputs it uses’ (p. 38) and excluding many other contractual relationships and firm activities. This incompleteness has not only tended to submerge his key idea – the comparison of the reduction in transacting cost of using market and the increase in coordinating cost of using the firm – but more importantly misdirected the attention of scholars who subscribe to his theory, especially, those people buying into Williamson’s asset specificity theory. What really concerns Coase is while he sees this weakness, other people who adhere to TCE do not seem to have detected this weakness.

Commenting on many of researches on the problems of the firm, Coase (1988: 42) points out many economists ‘seem to believe that vertical integration comes about mainly when there is asset-specificity’ (cf. Williamson, 1975, 1985), however, he argues ‘the conclusion…if not in error, is misleading’ (Coase, 1988: 43) because what decides whether to choose a vertical integration or an alternative governance structure, such as a long-term contract, ‘depends on the absolute relation [i.e., comparison] of the costs of these alternative arrangements’, not the degree of asset specificity whose relationship with vertical integration is doubtful (p. 43) and skeptical (p. 44) to Coase.

Having realized the weakness of his 1937 transaction cost theory and the consequence that many economists have tended to neglect the main activity of a firm, i.e., running a business, for instance, many economists tend to focus on some non-human-oriented factors, such as the texts of written contracts, Coase calls for developing a more comprehensive theory of the firm (Coase, 1988: 46) which may help us to understand firm activities such as vertical integration better. In addition, Coase acknowledges the interrelationships between the costs of transacting and the costs of organizing are extremely complex (p. 47), and in our present state of ignorance it will not be easy to discover (Coase, 1992: 718).
Another omission or flaw in Coase’s 1937 article is, as he himself realized (Coase, 1988: 47), 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.

3.2 Ghoshal and Moran’s (1996) call for an alternative theory

Having seriously criticized Williamson’s version of TCE being flawed, having limited explanatory power, and very bad for practice, Ghoshal and Moran (1996) call for developing a more attractive alternative theory to dislodge the dominant opportunism-based TCE theory. Although they were not capable of presenting one, they nevertheless have laid down some very important criteria and insightful suggestions for such a more attractive theory. There are six points in their recommendation: the new theory, firstly, should not deny the existence of opportunism; secondly, should not reject the need for control in organizations; thirdly, should not assume organizations are fundamentally inferior vis-à-vis markets, i.e., organizations have unique advantages over market (p. 31); fourthly, the purpose that provides the ultimate source of an organization’s advantage over the market must lie at the core of any theory; fifthly, the relationship and the quality of the relationship that determines the characteristics of the transactions must be taken into account (p. 37); last but not the least, the theory must adopt long-term efficiency as the criterion and address such variables as innovation, learning, and asset redeployability (p. 41).

In addition, Ghoshal and Moran (1996: 42) argue such a theory ‘is unlikely to emerge without considerable effort from strategy and organization scholars’ and cannot be a result of incremental modification of TCE because inclusion of some important variables such as ‘trust’ essentially falsify the basic tenets of TCE. Therefore, it is unlikely that the task of building a more attractive theory of the firm will be shouldered either by academic economists due to their language and tools at hand or by those strategy and organization scholars who have embraced TCE and propose incremental modification approach to making TCE better, or, more broadly, by any scholars who wholeheartedly support the ‘scientific model’ of social research (Bennis and O’Toole, 2005) which prescribes the best way to conduct social studies can be copied from natural sciences such as physics and chemistry.

3.3 We need an integrative, comprehensive and positive theory

I argue, there are some prevailing defects in the existing theories of the firm: being fragmented, incomplete, West-biased, and less human-oriented.

The theories of the firm are fragmented with each branches focusing mainly on one aspect of firm imperative. Transaction cost theories focus exclusively on how to reduce transaction cost, agency cost, cost of shirking, opportunum and underinvestment. Competence-based theories, such as dynamic capability view, knowledge-based view and resource-based view, put major emphasis on how to add value by building capabilities, knowledge, and resources. Stakeholder theory of the firm basically cares about how to satisfy the interests of all stakeholders involved in and affected by the operation of the firm. As Foss (1999) points out that some scholars have taken an ‘isolationist’ approach in conducting research in theory of the firm, namely, they believe each group of theories (i.e., economic theories and competence perspectives) can develop into full-blown strategic theory of the firm. I support Foss’s (1999) argument that an ‘integrationist’ approach is needed if we want
to develop an integrative and comprehensive theory of the firm. Since reducing costs, increasing value-added and distributing benefits generated by the firm are three inseparable imperatives of the firm, an integrative theory of the firm is needed (Casson, 2005; Foss, 1999) to deal with these three aspects of running the firm simultaneously.

Foss (1999: 727) argues that a ‘strategic theory of the firm’ needs to address four issues: besides the three Coasian themes, i.e., the existence, boundary and internal organization of the firm, one more issue to address is the competitive advantage of the firm, namely, which firm-specific factors account for superior rent-earning capability of the firm? Casson (2000) argues a modern theory of the firm should address the role of the entrepreneur, besides the three Coasian themes. He also argues introducing entrepreneurship into the theory of the firm is not a minor refinement, but represents a radical change toward an integrative framework which synthesizes other theories of the firm (Casson, 2005). However, I think all theorist of the firm from Coase (1937) onwards have neglected a very important question regarding the nature of the firm: how to ensure the firm generate economic rent and in an efficient way? Coase (1988: 47) expressed his regret that he did not examine what factors make one firm organizing same activities less costly than another firm in his 1937 article, which has consequently induced all his followers to believe the most important issues to be addressed by any economic theory of the firm are the three themes laid down in his original 1937 paper (cf. Foss, 1998:7). Still, I have one more puzzle: why all the theories of the firm do not talk about the role of employees (except discussing executive incentive pay) given the fact, as Simon (1995: 273) rightly points out that ‘counted by the head, most of the actors in a modern economy are employees’. So, I argue, a complete theory of the firm must be able to address all the seven issues: the existence, boundary, internal organization and competitive advantage, entrepreneurs’ and employees’ roles, rent generation and efficiency of the firm.

Also, the existing theories have a bias toward Western management thinking and practice and therefore do not well correspond to some aspects of the management practice in East Asia, especially, the Confucian societies such as China, Japan, Korea, and overseas Chinese communities in Southeastern Asia. In Confucian societies, guanxi, i.e., personal connections among ordinary people and between business partners, is widely practiced where people care about mianzi (i.e., face), renqing (i.e., reciprocal favor), and hexie (i.e., harmony in relationship), etc. Some particular business arrangements, like doing business without a contract or a detailed contract in China and the organizational governance structure of keiretsu in Japan and chaebol in South Korea, can hardly be understood by Western management theories, especially by Williamson’s opportunism-based theory of asset specificity. Obviously, the small firms at the periphery that highly depend on the big firm at the core of the keiretsu structure have very high degree of asset specificity in their contractual relationships with the core firm, and according to asset specificity theory these peripheral firms won’t be willing to invest in relationship-specific investment unless they merge into the core firm or are acquired by the core firm. However, this is not the case in the very stable long-term relationships within the keiretsu which is a distinctive characteristic of Japanese economy in comparison with the American. So, I argue, the new theory of the firm must account for these uncovered phenomena.

Most importantly, the existing theories are less human-oriented, i.e., not focusing on human behaviors and the motivations behind. A classical definition of economics is given by Robbins (1935: 15) as ‘the science which studies human behavior as a relationship between ends and scarce means that have alternative uses’. Coase (1998: 72) shortens to a simple statement ‘it is the study of human behavior as a relationship’ and points out ‘these days economists are more likely to refer to
their subject as “the science of human choice” or talk about “an economic approach”. The consequence of such a shift from human behavior to human choice is that ‘economists think of themselves as having a box of tools but no subject matter’, as Coase (p. 73) subtly notices. Although all the economic theories are human-involved, their main focuses are very little human-oriented, say, instead of studying human relationships, many TCE scholars shifted their attention to the texts of written contracts, ironically, Coase has helped such a shift of attention to the written contracts by initiating a Centre for Research on Contracts and the Structure of Enterprise at University of Pittsburgh (Coase, 1991:719) and establishing a contracts library which is now a part of the CORI K-Base hosting almost 700,000 corporate contracts at University of Missouri (Sykuta, 2009). Although I believe this contracts library is a rich resource pool, I am very concerned if it is a magnet to attract and divert economists’ attentions from more valuable human-oriented researches. The reason is straightforward: there are different types of contracts – explicit and implicit or structural and psychological contracts – the implicit and psychological contracts are more complex and important to our human behaviors and relationships. On the other hand, although competence-based theories are relatively more human-involved than economic theories, e.g., knowledge generation and sharing, they do not offer a convincing logic why employees would be willing to cooperate to build firm-specific competence, knowledge and capability. Even if competence-based theorists argue there are benefits for employees (i.e., after joining the firm) to do so, they have not gone deep enough to discover the human motivation behind these particular cooperative behaviors in question, and more importantly, their theories cannot explain why people start the firm as cofounders or join the firm as employees in the first place. So, I argue, the new theory of the firm needs to bring human behavior back to the center and in a proper way.

Outside economics, the stakeholder theories show their comparative advantage by being highly human-oriented, i.e., focusing on distribution of benefits of the firm to its all legitimate stakeholders. Unfortunately, stakeholder theories are idealistic at best and naïve at worst by arguing the firm should simultaneously satisfy all its stakeholders involved in and affected by the operations of the firm. Hence, they are impractical for the real business management. This triggers me to think about the academics’ moral and professional responsibility when developing theories and making normative suggestions. I strongly believe academics should take into consideration the practical consequences of their theories to be proposed. If we can predict the negative consequence of our theories, we should be very careful and even completely abandon those theories if very seriously consequence may be caused by those theories, say, opportunism-based theories as Ghoshal and Moran (1996) have condemned. Although some extremely normative stakeholder theories might cause some practical difficulties, they are nevertheless positive in the sense they place the interests

---

I once had an entrepreneurship-as-surplus-value-maximization theory of the firm which I thought made some sense. But almost at the time I formulated the theory according to the three Coasian themes I have realized it was such a bad theory in the sense that its publication would cause very negative consequences. I thus decided to completely abandon it. I can show you a little bit how bad this theory appears to be. The starting story is a person happened to realize he had made some extra benefit from buying from someone and selling to another in the market. Economics calls the benefit realized from buying as consumer surplus while the benefit from selling as supplier surplus. The extra benefit from buying-and-then-selling is called profit. And the economic behavior of making extra benefit by buying-and-then-selling as arbitrage.

Due to his greedy human nature, this guy wanted earn as much as possible this type of extra benefit in the market. But he soon realized the market for arbitrage might not be big or it was costly to identify such arbitrage opportunities in the market. So, he realized it would be much easier for him to maximize the extra benefits. Then he started a firm and became an entrepreneur. The existence of the firm can be argued to provide the entrepreneur legitimate rights to maximize surplus value he can appropriate from his employees. I don’t want to go on the analysis of the boundary and internal organization issues, which is not difficult to make anyway.
of non-shareholder stakeholders with equal importance with those of shareholders. So I argue the new theory should take into consideration that it is very meaningful to add into the theory some positive and normative elements.

4 The relationship-based approach: the foundations

4.1 Assumption about Human nature

The relationship-based theory of the firm (R’BT) assumes human nature is multifaceted which consists two parts: the negative part and the positive part. The negative part is further divided into two parts: self-interest and opportunism. The positive part also has two subparts: cooperation and altruism, which some economists seem to recognize, e.g., Fehr and Gächter (2000). Therefore, Williamson’s assumption about human nature is very biased and bad because he solely focuses on the evil part of human nature – opportunism.

R’BT also assumes that human nature is not fixed but environment-sensitive. According to dissonance theory (Aronson, 1980; Festinger, 1957), there is a positive feedback relationship between human behavior and the modification of human attitude as a means of reducing dissonance (Ghoshal and Moran, 1996: 21). Since human behavior is often constrained by surrounding environment, human attitude is therefore modified by the external signal from the surrounding environment. These signals may include: how others surrounding you behave, how others treat you, and how others surrounding you talk about human nature, etc.

To relax a bit our assumption about human nature being environment-sensitive, even if we cannot be sure if the two parts (and their subparts) can/will change in relative size, we are confident to argue that human attitudes derived from the respective parts of human nature can/will change. In social psychology, there is a crowding theory which argues human’s intrinsic motivation may be crowded out by extrinsic motivation, such as monetary incentive, and under some circumstances, it is difficult to restore the crowded-out intrinsic motivation even the extrinsic motivation is removed (Deci and Ryan, 1985; Lane, 1991; Lepper and Greene, 1978). This crowding-out effect is acknowledged by other organization scholars like Osterloh and Frey (2000) and some economists such as Frey (1994) and Frey and Oberholzer-Gee (1997).

There is also a crowding-in effect which allows external intervention, a reward or regulation, to strengthen or ‘crowd in’ intrinsic motivation (Frey, 1994: 335). Fehr and Gächter (2000: 159) point out that ‘many people deviate from purely self-interested behavior in a reciprocal manner’, namely, ‘in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model’. Because of the propensity of reciprocal behavior, many social psychologists led by Martin Seligman, the then president of the American Psychological Association, have called for a positive psychology movement which is ‘concerned with what is right with people and building on that instead of trying to just fix what is wrong with people’ (Luthans and Youssef, 2004:151). Positive psychology has spurred two related movements that emphasize the positivity and strength of human being. The first is the positive organizational scholarship (POS), and the second the positive organizational behavior (POB) (e.g., Luthans, 2002a, 2002b). I argue organization and management studies in general urgently need a positive theory of the firm and management in order to escape from the ‘pathological spiraling relationship’ (cf. Enzle and Anderson, 1993; cited in Ghoshal and Moran, 1996: 25) or the ‘self-fulfilling prophecy’ of opportunism (Ghoshal and Moran, 1996: 21). The proposed R’BV is intended to be such a positive
theory of the firm by purposefully emphasizing the positive side of human nature while downplaying the negative side.

4.2 The concept of relationship

The concept of relationship in R’BT is a simple term for two categories of relationships involved in the firm: the first is the relationships between the owner(s) or their representatives (i.e., managers) and the stakeholders of the firm, which are studied extensively by the principal-agent theory; the second is the relationships between and among employees of the firm, which is what Klein (1983: 374) terms ‘employee-specific interrelationships’ – many largely unwritten contractual arrangements within the firm – which have important impact on a large element of organizational value creation (p. 374). We can understand the concept of relationship from four perspectives: contract view, game view, option view, and capital view.

Commercial relationships are based on contracts, explicit and implicit. Kay (1993) identifies three types of commercial contract: spot, classical and relational contracts. Many scholars from different disciplines have discussed the concept of relational contract, such as law scholar Macneil (1974, 1978, 1980), economists like Olson (1965) and Grossman and Hart (1983, 1986) and sociologists like Bradach and Eccles (1989), Fox (1974), Granovetter (1985), McAuley (1963), Thompson et al. (1991) and Zucker (1986). In the R’BT, the relationship is underpinned by two or three types of contracts in Kay’s typology simultaneously. Take the example of a firm’s relationships with its customers. When a customer buys a firm’s products from the market, the buying-selling is a spot contract. Meanwhile, regulations of consumer rights protection serve as the legal contract between the firm and its customers, although this is often not discussed during the buying-selling process. Moreover, there might be a potential relational contract between the firm and its customers in that, if the firm takes care of its customers, they may become loyal to the firm. Another example is the relationship between a firm and its employees. The employer-employee relationship involves the legal employment contract as well as a psychological contract. If the firm looks after its employees more than the employment contracts require, the employees will be more satisfied psychologically, hence they would be willing to put more effort in their jobs as reciprocity, the commitment, productivity and creativity of the workforce may thus be increased.

Economists and management theorists frequently use game theory to analyze economic or business relationships (Casson, 2000; Kay, 1993; McMillan, 1992; North, 1990). IBM’s founder Thomas J. Watson Sr. once spelled out that ‘business is a game, the greatest game in the world if you know how to play it’ (cited in McMillan, 1992). Kay (1993) uses game theory to explain the cooperation and coordination of business relationships. Many scholars point out that mutual trust is the cornerstone of long-term relationship (Fuller and Lewis, 2002; Jarillo, 1988; Johannisson, 1987; Kay, 1996; Lorenzoni and Ornati, 1988). From a game theoretical perspective, building relationship is actually a repeated game in which trust, reciprocity, and mutual gains are necessary for an ongoing cooperation. The Japanese corporate system has had a harmonious capital-labour relationship owing to the cultural emphasis on trust (Matsumoto, 1991). A long-term relationship can never be established if the benefit of engaging a relationship is only partial, instead of mutual. In a sense, people or organizations establish or join in relationships for the sake of gaining benefits from the relationships. Chen (2004) explains why a guanxi network is vital to Chinese people and points out that the favour exchange is the norm of building a guanxi or a ‘guanxi web’. Su and Littlefield (2001) analyze the favour-seeking and rent-seeking dimensions of guanxi and claim reciprocity is the core of a guanxi. Larson (1992) also suggests that reciprocity norm is one of
important factors underpinning the duration and stability of relationship.

The option concept was borrowed by strategy scholars to explain strategic management as ‘a process of organizational resource-investment choices, or option’ (Bowman and Hurry, 1993: 760) (e.g., Bowman and Hurry, 1987; Hurry, 1993; Hurry et al., 1992; Kester, 1984; Kogut, 1991; Myers, 1984; Sharp, 1991). In financial market, the development and use of option contracts were motivated by risk-averse people’s intuition of ‘keeping options open’ against the uncertainty of future (Bowman and Hurry, 1993: 761). It can be said that strategies emerge from investment in resources which generate options for future investments. In the R’BT, relationships can also be understood from the option perspective. When two parties establish a relationship, each side does not know how the other side will behave in the future. Due to risk-aversion consideration, both parties want to strengthen the relationship through investing in the relationship, resulting in a gradual accumulation of each party’s sunk cost of engaging in the relationship. This incremental investment in the established relationship serves as an effective mechanism to reduce the opportunism. Another advantage of building relationship is to retain the option for future opportunity deriving from the current relationship. This is evidenced in the Chinese practice of building ‘guanxi web’ – a network of personal relationships based on reciprocity (Luo, 1996). By investing in different relationships, they essentially keep their future options open. While the western approach of relationship building is intended as an option to avoid risk and contain opportunism, the Chinese guanxi network is used as an option for access to future opportunities. This demonstrates the complexity and casual ambiguity (Einhorn and Hogarth, 1986) of relationship options compared to contractual options (i.e., option contracts). In this sense, we can say that the relationship engagement is more sophisticated than other strategic options (Bowman and Hurry, 1987).

And finally, the concept of relationship can be analyzed from a contract viewpoint. the concept of capital has been expanded to cover four distinct types: economic capital such as cash and machinery, human capital such as knowledge and skills, social capital such as networks and trust, and positive psychological capital such as confidence, hope and optimism (Luthans and Youssef, 2004). While economic capital and human capital can be generated independently by individuals, social capital can only be captured from embedded resources in a social network (Lin, 2001), i.e., in relationship. Lazega and Pattison (2001) propose that social capital, as resources embedded in social network, includes information, colleagues’ goodwill, advice and sometime emotional support, as well as many other means that serve individual and collective ends. Therefore, the social capital can enhance mutual trust and cooperation and therefore strengthen the interdependence between two parties, leading to the increase in commitment, productivity and creativity of individual human capital. The concept of relationship in R’BT includes social capital as well as the positive psychological capital because the development of a person’s confidence, hope, optimism and resiliency (i.e., positive psychological capital) is unlikely if he or she works in a very discouraging and conflicting environment. So a harmonious relationship is a precondition for developing the positive psychological capital.

4.3 The concept of harmony in relationship

When it comes to describing the state of a relationship, two opposite words naturally spring to our minds: conflict and harmony (Cheng, 2006). Conflict is normally associated with competition while harmony with cooperation. Scholars have often called for cooperation (Child et al., 2005; Doz and Hamel, 1998), especially, in the firm-stakeholder relationships (Gulati and Kletter, 2005; Jones,
Harmony in relationship is viewed as crucial in the Chinese Confucian societies and is placed in the core of some indigenous Chinese management theories, e.g., Xi, Han, and Ge (2006). Obviously, harmonious relationships can effectively reduce some contracting costs associated with agency, shirking, underinvestment and opportunism due to enhanced trust and cooperation (Jones, 1995). Therefore, harmony rather than conflict of relationships are the enduring goal to pursue for the majority of human beings.

Jones (1995) and Jones and Wicks (1999) point out trust and cooperation are fundamentally important in managing stakeholder relationships. I argue, besides trust and cooperation, how both parties of the relationship perceive whether their respective goals fit each other is also very important for the ongoing relationship (Ghoshal and Moran, 1996: 36). Accordingly, I dimensionize the concept of harmony in relationship, i.e., the quality of relationship (Ghoshal and Moran, 1996: 37), with three variables: degree of perceived trust, degree of cooperation, and degree of perceived fit between organizational goal and personal goal.

Normally, due to conflict of interests, in the case of employer-employee relationship as the conflict between organizational goal and individual goal, so, in order to achieve harmony in relationship, two parties of the relationship have to work together (i.e., cooperate mutually). For the management part, it is very important to take into consideration of the employees’ individual goals and try to accommodate them above the bottom line, i.e., ensuring the basic organizational goal must be achieved. Here, I use the term ‘basic organizational goal’ to indicate entrepreneur-owner or managers should be willing to compromise the so-claimed profit or shareholder value maximization expectation in order to leave room for realization of employees’ individual goals. So, managers should encourage employees to articulate their individual goals and help them to realize these individual goals. For the employees’ part, it is very important to internalize organizational goal so as to reduce the conflict of interests between organizational and individual goals. In this sense, if the Westerners values individualism and the Japanese value collectivism, I would argue the Chinese value individualism within collectivism, i.e., to realize self or personal goals within the collective. This is also the R’BT values.

4.4 Definition of the firm

The R’BT defines the firm as a nexus of two particular types of human relationship, the first is relationship between two persons who share a similar or same goal and are willing to partner to make that goal realized, which I term as zhitong-daohe relationship. Here, the Chinese word ‘zhitong’ means same-goal and ‘daohe’ means same-route. In a firm, such zhitong-daohe relationships only exist between cofounders (or owners) of the firm who share the same goal and work together toward that goal; the second type is the relationship between two persons who though do not share a same goal but nevertheless are willing to work together to realize their respective goals, which I term as zhibutong-daohe relationship. There is an extra word ‘bu’ meaning ‘not’ inserted into the word ‘zhitong-daohe’. So zhibutong-daohe means ‘not same goal still same route’. In a firm, such relationships exist between owner(s) and employees of the firm as well as among

---

5 Simon (1964:13) explains the employee’s internalization of organizational goal as follows: ‘the ability of a single individual to shift from one role to another as a function of the environment in which he finds himself thus helps explain the extent to which organizational goals become internalized…by whatever means the individual was originally motivated to adopt the role in the first place, the goals and constraints appropriate to the role become a part of the decision-making program, stored in his memory, that defines his role behavior’.
employees. To better understand this, we can imagine a story that there were thousands of British people who wanted to go to the New Land and there was the Atlantic Ocean to cross. The only feasible way was via ships because swimming to the other shore is impossible. Some people decided to own ships and run the transporting business to serve most of the others who wanted to go to the America with different dreams. There were many ships but some people must choose a same one and cooperate with each other in order to realize whatever their individual dreams even though they did not share the same goal with the owners and operators of that ship. Simply, the ship means the same route, just like the firm.

The treatment of the firm as nexus-of-relationships indicates that R’BT also subscribes to the view that the firm is just a ‘legal fiction’ (Jensen and Meckling, 1976) or a community where the owners of different input factors voluntarily come together based on at least daohe (i.e., same route) relationships. The owners of input factors include the entrepreneur as the owner of the business ideas, the creditors as the owners of financial capital, the employees as the owners of human capital and labor, and the suppliers as the owners of materials and components. As a community of voluntary memberships, the firm does not absolutely control anything because once the owners of inputs are fed up with the relationships they can choose leaving the firm. As a legal fiction, the firm does not absolutely own anything because any so-called firm’s own assets and capitals are merely undistributed surplus value appropriated or exploited from the co-creators of such surplus value. Simply, if every owner of input factor were fully compensated for his service, there would not have been any firm-owned assets.

Given the above definition, it is natural to argue the purpose of the firm is to increase the chances for all participants to realize their personal goals for its cofounders rather than to maximize firm profit or shareholder value. Klein (1983: 367) has pointed out that ‘in the real world…complete, fully contingent, costlessly enforceable contracts do not exist’ and hence ‘pure firm rent-maximizing behavior on the part of the agent-manager will not occur’ (p. 368). Recall the story again: the passengers chose the same ship to realize their dreams, not to help the owners of the ship to make more money. The firm can increase its members’ chances for realizing their individual goals because the potential synergy of working in a team means the chance for team success is greater than individual chance. Again, the chance for success of crossing the ocean by single-handed rowing a boat is slim if not impossible.

One of the functions of the firm can be argued as attracting and selecting the right people who can establish and maintain at least one type of the two particular relationships above mentioned: zhitong-daohe or zhibutong-daohe relationships. People may be attracted to (start) the firm as long as they see the chances for realizing their personal goals can be increased more than working somewhere else or working alone. So they join the firm because of being at least ‘daohe’, i.e., choosing the same route. But if they later find out the route cannot be the same, they may choose to quit the firm, even for cofounders. However, more people may be attracted to the firm. So the firm will attract and select people who can really share the same goal or the same route.

5 The relationship-based theory of the nature of the firm

5.1 The central thesis of R’BT

Let me first of all very briefly explain the fundamental difference between my relationship-based theory and the transaction cost theory. Transaction cost logic argues we live in a business world
where any type of transaction, no matter carried out in the market or within the firm, involves costs of transacting; and due to the presence of transaction costs, the rational choices in management and economy are to select the better mode of governance, i.e., the market, or the firm, or any hybrids in between, under which the total transaction costs are smallest, and the existence of the firm in the market is to reduce total transaction costs.

In sharp contrast, the R’BT places human relationship at the center of its whole arguments and takes harmony in relationship as the unit of analysis. R’BT makes an assumption that human beings prefer harmony to conflict in their relationships in human society. The reality is that we live in a world where disharmony often exists due to the conflict of interests in many human relationships and market-based relationships in particular. By market-based relationships, I mean the arm-length relationships different from more close relationships like between relatives and friends. Disharmony may exist in market-based relationships due to high costs associated with opportunism, distrust, and transacting, or the gap between the performance and expectations, or feeling isolated without true friends, etc.

So, an entrepreneur has the need of harmonizing relationships in order to reduce disharmony. The best solution is to form a firm in order to (1) bind zhitong-daohe (i.e., same goal and same route) people together to crystallize their shared goal and urge them work together in order to realize that goal; (2) attract people who might have different goals but nevertheless feel working together in this firm may increase their chances for success; (3) through the firm, they can also build a community for all the members of the firm (cofounders and employees) to socialize to fulfill their needs for socialization (Maslow, 1943). So R’BT argues the purpose of the firm is to maximize the perceived chances for success for all its members (owners and employees) by harmonizing relationships among them. However, it is not always the case that the firm-based relationships will be better than the market-based. The Chinese philosophy goes whenever benefits are generated by a new initiative; there must be some costs to incur at least latently. So, we cannot say the firm is always better than the market in dealing with relationships. People often say distance makes beauty, which demonstrates the argument that internalized intra-firm relationships may become disharmonious due to discipline squeezing personal freedom, frustration derived from internal competition for career, and feeling inequality due to power distance, etc. so there is no absolute benchmark which can best harmonize the relationships. It is contingent on how the firm really handle each relationship in question. The skill of relationship harmonizing is then absolutely a key people-specific or firm-specific capability which gives the person or the firm the competitive edge than others.

5.2 The existence of the firm

R’BT argues that the existence of the firm is due to the fact that while many kinds of relationships can exist in the market, like relationships between relatives, friends, schoolmates, etc., but there is no market for both of the two particular relationships aforementioned: ‘same goal and same route’ and ‘not same goal yet same route’ types. Simply, the same route means the firm. Put differently, such two relationships cannot be market-based in nature.

Say, a person called entrepreneur one day had an idea which he believed very good and he wanted to make that idea realized. He might feel he alone was not possible to realize that idea. So he needed to find others to help him. He went to the market and paid for some people to help him. However, those people who trade with him for monetary purpose might not share and care his idea per se – the services provided are just market exchanges – the outcome of these services might fall
short of the entrepreneur’s expectations. So he realized the pure market-exchange relationships could not increase his chance for realizing the idea. Then he thought about forming a long-term contract, e.g., a strategic alliance, with those people and organizations who might recognize the value of his idea, however, this hybrid governance structure, although it can to some extent increase his chance, would not be possible to maximize the chance for success of realizing his idea due to the risk-aversion on the part of the alliance partners in some circumstances (recall GHM’s underinvestment problem). So, there is only one option left – to form a firm with those who share and care his idea and then employ people to work directly towards realizing that entrepreneurial idea. Only under the hierarchy, i.e., the firm, the cofounders will take the risk and strive to realize that shared idea thereby their relationships are strengthened and more harmonized, the perceived chance for realizing the shared goal is not only increased but maximized.

After the firm is established, other people may be attracted to it if they feel this firm offers them a better chance to realize their personal goals, such as salary, career, and contribution to that industry, etc., than otherwise they may have in other places or working alone. As analyzed before, people may later on feel the chance to realize their personal goals is not up to their expectation after joining (or setting up) the firm, so they may quit. This is then the evolution process of the firm behind which is the dynamic evolution of the relationships between the members of the firm. In essence, the relationship and the perception of chance for goal realization are interdependent though not the same things. If the firm manages to establish harmonious relationships with its members, then the members tend to perceive better chances for realizing their goals, otherwise, if most of the members feel disharmony in their relationships with the firm (i.e., with managers and colleagues), they may quit and therefore the firm may fail.

5.3 The boundary of the firm

R’BT argues the expansion of the firm is determined by the relationship as well. The firm tends to expand as long as it can attract more people and maintain a certain degree of harmony in their relationships as well as provide better chances for newcomers’ goal realization. But the firm won’t expand infinitely because existing relationships may be disharmonized with the gradual expansion of the firm for different reasons. For example, many long-time employees may feel they are less cared about by the central management if they used to enjoy much more close relationships with the management. Another reason might be more people means less benefit can be shared if the profit growth is slower than the payroll growth. So, the firm tends to expand until the number of insiders quitting the firm due to perceived decreasing harmony by expansion is equal to or greater than the number of recruited with a perception that this firm offers them a better chance for realizing their personal goals than anywhere else or working alone.

As people may quit if they feel the chance for realizing their personal goal (i.e., disharmony in relationship) is not up to their expectation, this logic can also be applied to group of peoples (i.e., subunit within the firm). If the top management fails to maintain harmonious relationships with its subunits, the subunits may quit via management buy-out (MBO) or with other solutions. On the other hand, if the central management perceives there is no synergy or even disharmony between a subunit and the rest of the firm, the firm may divest it. Understandably, the divesture of subunits can be done for those units purchased from market, which means the firm might misjudge the potentials for enhancing harmony in the relationships by shifting the arm-length relationships into intra-firm relationships, or even if the initial judgment was not a mistake by bringing them into the firm, the switched relationships turned out to be less harmonious than prior arm-length relationships.
With regard to vertical integration, it is still the relationship between the firm and its business partner (another firm) that dictates the acceptable governance structure. Say, the firm has so far had a harmonious relationship with one supplier. If there is no any disharmonious contingency emerging in their relationship, they will carry out their businesses as usual. Imagine, if one day, the supplier suddenly requests the firm to pay an extra fee for the same transaction carried out before. Then whether this request will bring disharmony or not into their relationship will depend on how the firm perceive this request. There are several possibilities: (1) firstly, if the firm perceives this is fair because the supplier has had an increased cost for providing the same product or service due to the increased costs of raw materials, then in order to maintain a harmonious relationship, the firm will fully accept this extra fee request; secondly, if the firm perceive the request is not fair because the extra fee requested is higher than the increased cost to the supplier, then there are two options, (2) for one, if the extra fee is acceptable, the firm may tolerate and accept this request for the sake of maintaining the relationship though a bit disharmony is built in; for another, if the extra fee is not acceptable, then the decision of the firm depends on how important this buying-supplying relationship to the firm: (3) if it is not important and if it is possible to find another supplier for the same product and service, the firm will choose switch to another supplier; (4) if the relationship is very important or there is no other suppliers for the same product and service, the firm may purchase this supplier.

5.4 The roles of entrepreneurs, executives and employees

It is not difficult to argue that although the entrepreneur first has had an entrepreneurial idea he was determined to realize it, the realization of his idea would not be possible if there were no people (i.e., later managers and employees) willing to help him. In this sense, it is the entrepreneur who needs employees to help him. On the other hand, it is also intuitive that by conceiving that entrepreneurial idea and setting up the firm, the entrepreneur offered chances for employees to realize their individual goals. In this sense, it is the employees who need the entrepreneur (i.e., the firm) to help them. Therefore, it is natural to say the relationship between the entrepreneur and the employees is interdependent and mutual-beneficial.

According to standard economics theory, in a closed system of economy, if every factor is fully paid for its contribution, there is no rent possible to be generated by any productive organization (e.g., the firm). But in reality, we are familiar with the notion of firm-owned capital. From a Marxist perspective, this firm-owned capital is in fact accumulated undistributed surplus value appropriated by the firm from its input-factor providers, the employees in particular. In most employer-employee relationships, it is the employer (the firm) who is able to appropriate surplus value from its employees (i.e., human capital providers) due to bargaining power asymmetry. In management theory and practice, we often encounter discussing executive incentive pay and why it is important. It is correct to say executives are very important because they are the decision makers whose intentions and actions will definitely have a big impact on the overall performance of the firm. Nevertheless, I argue we might have gone too far to over-attend executives while downplaying the importance of the majority of the people involved in and serving the firm – the employees. If the intuition tells us that the executives are often overpaid (see the debate on whether US CEOs are overpaid between Kaplan (2008a, 2008b) and Bogle (2008) and Walsh (2008)), the corollary is that employees are often underpaid in order to finance overpaying executives, given the conventional expectation of shareholder value maximization.

So, R’BT argues the entrepreneur should think carefully about his role and the roles of the
executives and the vast majority employees. While individual executives are more important than individual employees, it would be blunt to think the few executives have made the biggest contribution to the performance of the firm as if the vast majority of employees are just tools used by the executives to complete tasks designated. R‘BT supports some scholars’ view that the most important resources of the firm are the employees who collectively really make a difference.

Here it is time to clarify the difference between the entrepreneur and the non-entrepreneur employees. R‘BT does not subscribe to those theories which conceptualize the entrepreneur as a special type of person who are genius, born innovator, cleverer than others, and charismatic, etc. I argue there are only two distinguishing characters really separating the entrepreneur from non-entrepreneur. Firstly, entrepreneurs are risk-taking under uncertainty while non-entrepreneurs are risk-aversive. Secondly, an entrepreneur is unwilling to surrender his goal to others’ goal, i.e., his goal is the priority, except as an expedient. A non-entrepreneur is more willing to accept realizing his goal depending on the realization of others’ goals, i.e., being prioritized by others’ goals. It is no doubt that some of the entrepreneurs may be like that, but they are nevertheless not the distinguishing characteristic of entrepreneur. Here is one example. Mr. Liu Bang, the founding emperor of China’s Han Dynasty, once admitted he was not as good as Mr. Zhang Liang in strategy, not as good as Mr. Xiao He in organization and not as good as Mr. Han Xin in fighting a war. Also, he was not clever if judged by his career achievement before he started his entrepreneurial career, i.e., overthrowing the ruling emperor; he was not charismatic given the story that he once wanted to forsake his kids in order to escape. Even, he used to be greedy and lecherous. In many ways, he was not the material for emperor, but he succeeded because he was risk-taking when he saw the time was ripe while the other three brilliant guys only wanted to get a job by helping others to win. More important, Liu Bang himself pointed out that he has one particular advantage which is he could use these brilliant guys and used them well. There were many stories how well Liu Bang managed his relationship with those brilliant guys who come to help him.

In reality, many people may have very good ideas but they are risk-aversive, so they dare not to take risk: (1) they may be too much bothered by failure and opportunity costs; (2) they may have no sufficient capital to venture a business to realize his idea, and due to risk-aversion, they dare not to take loans or borrow from relatives and friends. So even they may have entrepreneurial spirit, they cannot become a real entrepreneur. A real entrepreneur may set up a firm with co-founders because of zhitong-daohe relationships, or starts a single-owner firm because either he cannot identify any zhitong-daohe partners or he does not want to share ownership of the firm with a co-founder. If the single-owner firm is not perceived to be able to offer a greater chance for other people’s goal realization than other places, this single-owner firm may stay as sole proprietorship.

5.5 Internal organization of the firm

R‘BT offers another viewpoint on authority in the firm. Rather than lean towards either Coasian fiat-type authority of the employer or Alchian and Demsetz’s no authority at all in the firm, R‘BT argues the owners or their representative managers do have authority over their employees or subordinates. But the 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 to, under the precondition of harmonious relationship, prioritize its
organizational goals over its employees’ individual goals.

This prioritization-under-harmonization capacity based authority indicates that if the employer abases its bargaining power, 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 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.

R’BT also does not prefer the formal control argued for by TCE scholars, especially Williamson. In China where relationships and harmony are highly appreciated and cultivated the management philosophy goes it is not impossible to achieve a perfect type of control: governing without intervening. Governance without interference can be understood like this: 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 each level follow this ‘governing without interference’ principle to manage their subordinates who manage their tasks, then the whole company can be governed without interference.

5.6 Competitive advantage for rent generation

An economic organization often strives to make economic rent, i.e., above average profit. There are four major types of economic rent: first, Porterian rent derived from high bargaining power in contractual relationships. According to Porter’s (1980) model of five forces of industry, we understand that a firm can squeeze or appropriate the value-added of its suppliers and customers, and its employees as well. Obviously, this rent is generated at some other’s expenses; second, Ricardian rent derived from superior resources the firm control, which is argued for by the resource-based view of the firm; third, Schumpeterian, rent derived from entrepreneurial innovations or creative destructions (Schumpeter, 1934) which generate more value than existing economic system. The last one, I term as Confucian rent, which is generated when under harmonious relationship people work in a more committed, more productive and more creative way than otherwise if there is no harmony in the relationship. The extra output than usual by working more committedly, productively and creatively is called Confucian rent.

R’BT argues harmonious relationships provide the firm a relationship advantage, which can be argued as the foundational level competitive advantage, while the low cost and differentiation advantage are the end level competitive advantage. Any other competitive advantages outside these two categories are to be argued as intermediate level advantage, such as innovative, entrepreneurial and quality advantages. R’BT further argues there must be a transformational mechanism for the firm to transform the foundational relationship advantage through intermediate level advantages into the ultimately end level competitive advantage, i.e., low cost, or differentiation, or combination of the two.

5.7 Economic efficiency

Economics is concerned with efficiency (i.e., best, optimal) use of limited resources (Marschak and Radner, 1972: 3). However, many economists have paid more attention on economic resources such as physical and financial resources than on human resources. R’BT argues human resources of an individual such as time, effort, and knowledge are scarce resources to himself. Where to work, what
to work with, and how to work are very important decisions which will decide the efficiency of use of his human resources. Obviously, working in a disharmonious environment will hamper the allocation efficiency of human resources. So in order to improve the allocative efficiency, the human resources must find a harmonious work place.

But efficiency of allocation is not equal to efficiency of utilization of human resources. So, even within a harmonious working place, the employer should encourage the commitment, productivity and creativity of the employees, if so, utilization efficiency can be improved.

6. A candidate theory for organizational economy and management?

I argue the relationship-based theory has the potential to be a general theory of any types of organization, such as economic organizations like the firm, not-for-profit organizations, governmental agencies, social clubs, political parties, and even the nation because any human organizations are faced with the problem of conflict of interests which will impact on harmony in relationships depending on how these conflicting interests are handled; to reduce the conflict of interests, the best way is to harmonize relationships by mutually accommodating or internalizing the other party’s interests into its own consideration; the degree of harmony in the relationships will in turn determine the performance of the organization in question. The argument that R’BT can be adopted by any level of organization is demonstrated by the fact that Chinese government has officially adopted a relationship-harmonization strategy as it has announced to build a harmonious society internally and called for building a harmonious world internationally. Another example can be a political party. At its embryo stage, it is the zhitong-daohe relationships that attract and glue the cofounders together; their ambition is to realize an entrepreneurial idea which is impossible for any individual alone to realize. After the party is formed, these cofounders work together and strive to make it bigger by attracting people who may share the same ideas or may not share the same ideas but nevertheless feel this party offers them better chances for realizing their individual ambitions. This is especially for a revolutionary party whose ambition is to obtain the rights to regime. After the party succeeds and becomes a ruling party, more people would be attracted to it for all kinds of reasons: basically they just want to realize their goals within and with the ruling party. But a ruling party may fail if it cannot harmonize its internal relationships with its members and external relationships with its people due to the conflict of interests.

The R’BT can also be a candidate for ‘a theory for the organizational economy’ (Ghoshal and Moran, 1996: 37). As analyzed before, relationship-harmonization is a means to improve efficiency in resources allocation and utilization, especially for the most important human resources, i.e., time, effort, and knowledge of individuals. In the economics academia, many people tend to make distinction between capitalism and communism as two conflicting economic systems and taken for granted an assumption that capitalist economy is logically more efficient than communist economy. However, I argue, this is an illusion and a false assumption reinforced by the collapse of Soviet economy in late 1980s and early 1990s. It can be argued that the early Soviet Communist economy was extremely efficient and effective given the undeniable fact that Soviet Russia quickly industrialized itself and achieved its ‘catch-up’ ambition before the Second World War. But why and how can such a centrally planned economy in which price mechanism is suppressed achieve such high growth rate? The answer lies in human relationships and the harmony in relationships within an economic system. If human, the most important resources in economic system, enjoy harmonious relationships, for reasons like having been given the rights to personal freedom, to work, to productive means, and whatever reasons, they might be motivated to work committedly,
productively, and creatively, then the economic performance as a whole will be largely improved. This has been the case for the early success of Soviet economic development. Unfortunately, Soviet leaders and the Soviet Communist Party as a whole did not handle the internal and external relationships harmoniously, so, the efficiency of its economy was dampened and disharmony within its people was intensified and turned into frustration and intolerability. As a sharp contrast, the Nordic countries have adopted a distinctive welfare state model, in which people enjoy more equality than other capitalist countries like USA while more freedom to economic matters than Soviet Communism, this Nordic model of economic system is working well. I argue the reason might be the Nordic countries have managed human relationships relatively speaking more harmonious than any other society.

Total relationship management

This relationship-harmonization theory can be used as a theory of organizational management as well. The Chinese management wisdom tells us management can be simply said as managing people and managing tasks. As tasks are performed by people, the key to effective management is managing people well and motivate them to perform tasks well. Xi (1989) first proposed and developed a HeXie Theory (harmony theory). He divides the Chinese word HeXie (i.e., harmony) into two words: He and Xie, with the former meaning harmony and the latter meaning coordination. He then propose a dual-principle management theory, namely, under more dynamic and uncertain situation, we should rely on He-Principle which values making system and subsystems more harmonious by informal and artful means, while in more certain and static situation, we can use Xie-Principle which emphasizes coordination under explicit rules, regulations and routines.

What I disagree with his dual-principle is I instead argue these two principles are not parallel and the He-Principle should be the fundamental and foundational principle and Xie-Principle is a second-order yet important means: when it comes to management, we have to first of all use the He-Principle of building and harmonizing relationships with people we are working with, no matter your boss or your subordinates, or external business partners, then under the condition of having harmonious relationships within the workplace, we then use Xie-Principle to design effective coordinating and incentive systems in order to motivate people to work committedly, productively, and creatively. This is my version of HeXie management (harmonious management).

R’BV argues since relationship lies at the core of the theory of the firm and harmonious relationship provides the firm the basis for competitive advantage, then the essence of management of the firm or any organization should be managing relationships among the constituents of the organization. I term this as total relationship management (TRM). TRM can be explained by the model of relationship harmonization (see Figure 1), the rationale behind is harmony in relationship mainly depends on the degree of perceived fit between organizational goal and individual goal.

The first step – Consideration – who are the strategic stakeholders and what are their expectations? I think, even within one group of stakeholders, such as employees, there are some people who are more strategic at a particular time, so it is very important to consider them first and making relationships with them harmonious first. We can use the stakeholder typology developed by Mitchell, Agle and Wood (1997: 874) which maps out the identification and salience of stakeholder according to three relationship attributes: power, legitimacy and urgency.

The second step – Evaluation – the managers need to know whether the strategic stakeholders are
happy with their relationships with the management and their colleagues of the firm, i.e., to what extent they perceive their relationships are harmonious. This can be done through ‘conducting performance audits’ and ‘revealing gaps’ (Preble, 2005: 415). As discussed in section 3.2, harmony of relationship is determined by three dimensions: degree of perceived trust, degree of perceived fit between organizational goal and personal goal, and degree of cooperation, managers should help subordinates identify the real and potential factors which may harm their relationships. More importantly, managers should encourage subordinates to articulate their individual goals and try to accommodate them within the organizational goals.

The third step – Communication – managers should communicate two things with subordinates. The first thing is how the managers perceive the performance of the subordinates. Since harmony of relationship is desired from two parties of the relationship, it is a joint effort to make the relationship harmonious and therefore, each party should let the other party know and understand how and why they perceive the relationship. In conventional management, it is often the case that managers will tell the subordinates how they feel about their performance (and the relationship) but not the other way around. R’BV proposes a positive management should make the two-way communication happen and the management would better to let the employees to speak first.

The fourth step – Participation – management should let strategic stakeholders participate in decision-making process. Social scientists have long analyzed the importance of procedural as well as outcome justice of distribution of wealth (e.g., Deutsch, 1975; Greenberg, 1990:401; Philips et al., 2003: 490). Some scholars point out that ‘people are more apt to accept decisions and their consequences if they participated in making them’ therefore ‘participation in the decisions which affect one’s well-being helps to legitimize such decisions’ (Deutsch, 1975: 139).

The fifth step – Collaboration – management should collaborate with strategic stakeholders to solve the real and potential conflicts between the organizational goal and the individual goals of these stakeholders, or at least to design a conflict-solving mechanism with which in place stakeholders may tend to be more cooperative if such conflicts really emerge than were otherwise it absent. Another meaning of this step of collaboration can be that collaboration between the employees and management as well as among the employees is a natural outcome due to the prior efforts of relationship harmonization as indicated in the above four steps.

Since there are many stakeholders involved in and affected by the operation of the firm, according to normative stakeholder theory, each stakeholder’s interest should not be sacrificed in order to maximize another stakeholder’s benefit. As I argued before, this normative stakeholder is not practical by arguing that we should attend to all stakeholders’ interests simultaneously, then it is very natural to argue that the five-step model of relationship harmonization should be a circular one, namely, after each round, we should reconsider who else (stakeholders) we should include into our strategic consideration for harmonizing our relationship with them, and once identified, we should strive to do so (i.e., relationship harmonizing). Taking the stakeholder argument into account, we should be convinced that we will always have responsibility to expand our scope of attention. Therefore, this relationship harmonization is essentially a never-ending process.
Xin Li is a PhD candidate in Business Economics at Asia Research Centre, Copenhagen Business School. His current PhD research is about China’s comprehensive national power.

Acknowledgement

I want to thank Nicolai J. Foss and Peter G. Klein for their excellent teaching in a 3-day PhD course on the theory of the firm at Copenhagen Business School during 22-24 June 2009 where the author has learnt many insights of the economic theories of the firm.

Some of the ideas of this paper have been developed since 2006 when I was an MPhil/PhD candidate in Strategic Management at Royal Holloway, University of London. I have presented some of these ideas under the label of ‘a relationship-based view of the firm’ (R’BV) at annual conferences of the British Academy of Management in 2007, Strategic Management Society in 2007, Academy of Management in 2008 and European Business Ethics Network-UK in 2008. My idea of writing this paper was conceived on 20 June 2009 when I was reading the suggested literature in preparation for their course on the theory of the firm. When I read Foss & Klein’s (2005) paper about entrepreneur as judgment theory, especially when I read their argument that the firm exists because there is no market for judgment, I was suddenly enlightened how to reformulate my prior R’BV: I might better rearrange all my prior viewpoints around a central logic, namely, the firm exists because there is no market for some particular types of relationships which in nature are firm-
based. From 20 June on, I have kept on reading, thinking, and taking notes until 27 June 2007 I finally could not help starting my journey of writing down this paper. During the writing period, I still kept reading, thinking and correcting my writing. It has been an arduous yet exciting process of learning.

My special thanks go to Verner Worm who is Professor of Chinese Business Management scholar at Asia research center, Copenhagen Business School. It is Verner who has encouraged me to do some indigenous Chinese management research and who has awakened my lost interests in management research after I quit the School of Management, Royal Holloway, University of London (RHUL) where I had done 2 year PhD research in strategic management with a genuine and uncompromising ambition of developing a Chinese theory of strategic management which was not supported by anybody of my PhD semi-annual review committee there and my initial relationship-based view of strategy was seen as having potential for a Heathrow airport book rather than academic material for a PhD project, which I never agree. Nevertheless, I would like to thank RHUL for granting me a PhD scholarship for those 2 years.

Verner’s support for my doing management research has also been helpful for me to work out a linkage between management topic and my current PhD project, i.e., China’s comprehensive national power. The linkage I have identified is as follows. In agreement with existing literatures which normally argue comprehensive national power includes hard power and soft power, I nevertheless argue that there should be one extra category of national power – management power – because such things like national cohesion is a form of national power which do not fit well into the dichotomy of hard and soft power. National cohesion, or management power in general, is not hard power because its soft nature. However, Joseph Nye’s definition of soft power stipulates soft power is something other countries admire which give soft power a positive undertone. The management power of a nation is not necessarily always positive and admirable, say, the national cohesion and management power of Nazi Germany in 1930s. Therefore, a separate category of national power is needed to give a more complete picture of comprehensive national power.

My argument goes on that management power is dependent on how well matched between three elements: national work ethic, national management philosophy and national management practice. National work ethic to economic development is well argued by Max Weber (1904), Herman Kahn (1979), Hofsted and Bond (1988) and S. Gordon Redding (1993). National management philosophy is very important and can be crystallized as management theories. National management practice is also an important source of country-specific advantage, such as Japanese management system. However, I would argue, Japanese management system should be better understood as a unique Japanese management practice rather than a management theory or even management philosophy. It can be argued Japanese share very similar management philosophy with Chinese: the Confucianism.

Since Chinese economic development is catching the world’s attention, scholars in the West as well as in the East should carefully examine the uniqueness of indigenous Chinese management. There are some things special, like the practice of guanxi (relationships), emphasis on hexie (harmony) and worshipping wuweierzhi (governing without intervening). As a Chinese, I understand that there are so many good discussions and writings on Chinese management in China, however, due to all sorts of reasons, they have not made a substantial impact on the mainstream (the Western) management academia. As a Chinese scholar studying in the West, I am always conscious of the situation and the need to filling the void. So, this ‘relationship-based theory’ paper is just one of my efforts toward that goal.
Reference


Deutsch, M. 1975. ‘Equity, equality, and need: What determines which value will be used as the basis for distributive justice?’, *Journal of Social Issues*, 31(3): 137-149


Kaplan, Steven N. 2008a. ‘Are U.S. CEOs overpaid?’, Academy of Management Perspectives, 22(2): 5-20


Knight, Frank H. 1921. Risk, Uncertainty, and Profit. New York: Augustus M. Kelley


Luthans, Fred. 2002b. ‘Positive organizational behavior: Developing and managing psychological strengths’, *Academy of Management Executive*, 16(1): 57-74


