REASSESSING THE FUNDAMENTALS AND BEYOND:
RONALD COASE, THE TRANSACTION COST AND
RESOURCE-BASED THEORIES OF THE FIRM AND THE
INSTITUTIONAL STRUCTURE OF PRODUCTION

ANOOP MADHOK*
David Eccles School of Business, University of Utah, Salt Lake City, Utah, U.S.A. and
Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands

In this paper, three points are argued. The first is that Ronald Coase, best known as the forefather of transaction cost theory, foresaw many of the critical questions that proponents of the resource-based view are concerned with today. The second is that resource-based theory plays a potentially much more critical role in economic theory and in explaining the institutional structure of production than even many resource-based scholars recognize. The last point is that a more complete understanding of the organization of economic activity requires a greater sensitivity to the interdependence of production and exchange relations. The arguments presented in this paper highlight important, but relatively ignored, elements in Coase’s work that inform strategy research. More importantly, this paper makes the case for a triangular alignment between the triumvirate of governance structure, transaction, and resource attributes and demonstrates how the identity and strategy of a particular firm influences how its resources interact with the transaction and how the firm chooses to govern it. The general argument is then applied to the context of interfirm collaborative relations, where the key focus is broadened from just cost to also include skills/knowledge and the interdependence between cost and skills with respect to firm boundaries, both in terms of choice and nature. Such a broadening of focus enables us to additionally examine the transacting process as a productive endeavor, which underpins the co-evolution of the competencies of partner firms.

INTRODUCTION: THE ORGANIZATION OF ECONOMIC ACTIVITY

It has been over six decades since Ronald Coase wrote his now famous article ‘The Nature of the Firm’ (1937), in which he grappled with the nature of the firm within the context of the institutional structure of production. Departing from one of the fundamental tenets of neoclassical economics, he questioned the notion of frictionless markets and argued that there was ‘a cost of using the price mechanism. The most obvious cost of “organizing” production through the market mechanism is that of discovering what the relevant prices are’ (Coase, 1937: 390). These transaction costs (TC) make it more efficient to organize an activity within the institution of the firm.

Coase’s main purpose was to explain why economic activity was organized within firms. It was not his purpose to predict which particular transactions would be organized within the firm. Williamson’s (1975, 1985) major contribution was to make the theory more predictive by approaching the firm as a governance structure more microanalytically and by identifying the particular transaction characteristics, in particular asset specificity, that play an important role in comparative institutional assessment. Since Williamson, the theory has shifted away from Coase’s initial and
more general treatment to a concern with issues of appropriation, ownership, alignment of incentives, and self-interest that characterize Williamson’s influential work. Yet, in spite of their differences in emphasis, both Coase and Williamson saw firms and markets as alternate means of coordination, the firm being characterized by coordination through authority relations and the market being characterized by coordination through the price mechanism (Coase, 1937; Williamson, 1991a).

TC arguments have, however, not gone unchallenged. A growing body of work in strategic management, collectively labeled the resource or capability-based view of the firm, contends that the reason an activity is conducted within the firm is not market failure (i.e., the cost of transacting through the market) but rather firm success: the firm as an institution enjoys an ‘organizational advantage’ which enables it to organize economic activity in a manner that markets simply cannot (e.g., Teece, Pisano, and Shuen, 1997; Kogut and Zander, 1996; Ghoshal and Moran, 1996; Conner, 1991; Madhok, 1996, 1997). Teece’s work in particular (Teece, 1986, 1990; Teece and Pisano, 1994; Dosi, Teece, and Winter 1992, among others), while building on Williamson’s, goes a step beyond it by addressing not just efficient contracting and TC, but also production and organizational economies and ‘the distinctive ways that things are accomplished within the enterprise’ (Teece, Pisano, and Shuen, 1997: 528) as a result of firm routines (Nelson and Winter, 1982). These distinctive ways by which firms manage their resources and capabilities can result in superior performance and function as an enduring source of competitive advantage (Barney, 1991; Peteraf, 1993).

In short, two important and popular approaches to understanding the firm within the context of the institutional structure of production are the transaction costs (TC) and the resource-based (RB) views. TC scholars have focused mainly on the role of efficient governance in explaining firms as an institution for organizing economic activity, while RB scholars have tended to emphasize the role of competitive advantage (Barney, 1991; Conner, 1991; Peteraf, 1993). Good reasons exist for both these analytical strategies and much important work has been produced by researchers from both traditions. Yet, most people who are familiar with Coase’s (1937) landmark work ‘The Nature of the Firm’ associate it almost exclusively with the question of why firms exist, which sets the agenda for TC economics.

In this paper, three points are argued. The first is that Coase anticipated many of the critical issues that RB scholars are concerned with today, including the central question of performance differences among firms. The second is that RB theory plays a potentially much more critical role in economic theory and in explaining the institutional structure of production than even many RB scholars recognize. The last point is that a more complete understanding of the organization of economic activity requires a greater sensitivity to the interdependence of production and exchange relations, a point more recently emphasized by Coase (1988).

The paper is organized as follows. In the following section, Coase’s contribution to the RB view, hitherto not fully recognized, is discussed. This is followed by a discussion of the interrelationships between production and exchange at a general level of abstraction. The case is made for a triangular alignment between the triumvirate of governance structure, transaction, and resource attributes and it is demonstrated how the identity and strategy of a particular firm influence how the firm’s resources interact with the transaction and how the firm chooses to govern it. In the fourth section, the general argument is applied to the context of interfirrm collaborative relations, where the key focus is broadened from just cost to also include skills/knowledge and the interdependence between cost and skills with respect to firm boundaries, both in terms of choice and nature. Such a broadening of focus enables us to additionally examine the transacting process as a productive

---

1 Since most of the work within the TC framework has been modeled after Williamson, the criticism has been largely directed at the arguments set forth by him.

2 Resources, broadly defined, have often been used in the literature in a generic sense to also include capabilities (e.g., Barney, 1991). Others (e.g., Teece et al., 1997) claim that capabilities refer to how firms manage resources. Moreover, capabilities and their underlying routines can be considered as forms of knowledge about how to carry out productive tasks (Langlois and Robertson, 1995; Kogut and Zander, 1992). Although one can get into detailed discussion of semantics about the differences between the various terms, ultimately they are all interested in the similar question of performance differences between firms. For the purpose of this paper, I use the term ‘resource-based’ in an encompassing manner to denote the collective technological, organizational, and ultimately commercial capacity of the firm.

3 In order to elucidate and illustrate my arguments, I refer to, and draw from, Coase’s various works throughout this section of the paper.
endeavor, which underpins the coevolution of the competencies of partner firms. Some concluding remarks are presented in the fifth and final section.

RONALD COASE AND THE RESOURCE-BASED VIEW

There are two principal questions with respect to the organization of economic activity. The first is ‘Why is an activity organized within firms and not purchased through the market?’ The other, equally important, question is ‘Why is an activity organized within a particular firm (or firms) and not any other?’ That is, how is economic activity distributed among firms? The main difference between the two questions is that the first relates to the institutional level, and is accordingly concerned with the institution of firms and markets as a whole, while the second relates to the firm level, and is accordingly concerned with specific firms.

In raising the question ‘Why is an activity organized within firms?’ or, putting it differently, ‘Why do firms do what they do?’, Coase also asked another question ‘Why is all economic activity not organized within one large firm?’ or, putting it differently again, ‘Why do firms not do what they do not do?’ This second question, like the first, is a deceptively simple one, but one which has considerable import. At its heart lies the notion of competitive advantage and disadvantage, the very issue that RB scholars are most concerned with. If the most obvious cost of choosing to make rather than buy is that of discovering the relevant prices, then the less obvious cost is that of choosing to make something which a firm is inadequately equipped to do competitively. In Coase’s words, at a certain point:

[T]he entrepreneur fails to place the factors of production in the uses where their value is greatest, that is, fails to make the best use of the factors of production. Again, a point must be reached where the loss through the waste of resources is equal to the marketing costs of the exchange transaction in the open market or to the loss if the transaction was organized by another entrepreneur. (Coase, 1937: 394; emphasis added)

The argument explicitly recognizes the importance of differences in production costs, with the comparison being essentially a three-way one in which, for an activity to be organized within a firm, the costs of doing so would need to be (a) not only lower than that through the market, but (b) also lower than that within any other firm; otherwise, it would be more advantageous for the other firm to organize it.

These two comparisons are not unrelated. Considering that the bulk of economic activity is carried out by firms (Coase, 1937; Simon, 1991), the market, for the most part, is an implicit and abstract representation of other firms (Demsetz, 1988; Chandler, 1992; Coase, 1937; Simon 1991). To underscore this point, Coase elaborated in his 1987 lectures on what he meant in his original paper:

What emerges from this interfirm competition . . . is a situation in which, apart from the purchase of the services of factors of production and retail trade, most market transactions will be interfirm exchanges. (Coase, 1988: 40)

Effectively then, market exchange mostly amounts to exchange between firms. That is, the division of labor between the firm and the market is actually the division of labor between firms, and therefore has to do with the distribution of economic activity between firms. This very issue, framed in terms of performance differences across firms, is what the RB view of the firm is fundamentally concerned with.

The general argument on firm boundaries also extends to prominent areas of application of RB theory such as diversification and firm scope (Penrose, 1959; Ramanujam and Vardarajan, 1989; Mahoney and Pandian, 1992). The RB view cautions firms to broadly ‘stick to their knitting,’ with the decision with respect to in-house production hinging on capability-related considerations to do with routines. Routines provide the basic building block to the RB line of argument and explain various phenomena such as firm inertia, learning, path dependence, knowledge stocks and flows, and other like notions (Dosi et al., 1992). Since each firm has a basic area of competence, gradually accumulated through experience, this becomes the source of its competitive advantage as well as a competitive constraint. Overextension of its activities into domains which are too diverse and dissimilar not only dilutes the strength of its competence, but also increases the costs of organizing in-house due to the lack of experience and expertise in these areas. Such behavior is therefore ill advised. On the other hand, closely related activities economize on
costs since resources and routines can be lever-aged across them. The argument emphasizes the effect of focus, in the absence of which a particular resource tends to lose value when unintelligently transferred to arenas that are too different.

While the logic is both theoretically and intuitively appealing and enjoys popular currency, Coase anticipated the very same argument both in his original paper:

> It would appear that the costs of organizing and the losses through mistakes will increase with an increase in ... the dissimilarity of the transactions. (Coase, 1937: 397)

As well as in subsequent ones:

>[T]he costs of organizing an activity within any given firm depends on what other activities it is engaged in. A given set of activities will facilitate the carrying out of some activities but hinder the performance of others. It is these relationships which undermine the actual organization of production. (Coase, 1972: 64)

But this [i.e., the question of why there are firms] does not tell us what the institutional structure of production should be. That depends on which firms can carry out this particular activity at the lowest cost and this is presumably largely determined by the other activities that the firms have undertaken. (Coase, 1988: 40)

Basically, with his introduction of TC into the comparative institutional assessment, Coase departed from neoclassical economic orthodoxy’s exclusive focus on the firm as a production function. Yet, by emphasizing TC and by arguing that one firm will internalize the activities of another only if the TC saved is greater than any additional production costs, he in essence superimposed TC (the exchange dimension) onto production costs (the production dimension). Building upon Coase, Williamson basically did the same. While stating that the mode of organization depends on the sum of production costs and TC, Williamson treated production in a somewhat ‘stylized’ manner, holding production technology constant and exogenously determined (i.e., equally available to all firms) as a pedagogical device to concentrate on his exposition of TC. The comparative institutional choice, therefore, effectively rests on a transaction costs assessment (Williamson 1985: 89).

In contrast to TC theory’s departure from neoclassical orthodoxy, as a result of its analytical focus on friction in exchange, RB theory departs from neoclassical theory as a result of its analytical focus on friction in production. The production function is viewed not merely as a technical input–output transformation function, freely available to all firms and taken as a given, but a more nuanced and sophisticated one in which the technical component and the organizational component are intimately fused together. That is, the production function becomes partially endogenous, with a firm’s ‘organizing technology’—as defined by its routines—playing an important role in the transformation of inputs into outputs. The ‘organizing technology’ denotes that, even if two firms had access to similar inputs and technology, there could still be a difference in performance due to differences in organizing skills and abilities. This is where issues like a firm’s activity portfolio, past experience, inertia, learning, path dependence, knowledge stocks and flows, etc., become central.

While both theories are important departures from key tenets of neoclassical economics, Williamson’s landmark work ironically resulted in a subsequent underemphasis on production costs and detracted from a more full-bodied understanding of these costs, a situation currently in the process of correction by the RB view. On this point, the following statement by Coase half a century later is telling:

>[I]f one is concerned with the further development of the analysis of the firm’s activities, the way in which I presented my ideas has, I believe, led to or encouraged an undue emphasis on the role of the firm as a purchaser of the services of factors of production and on the choice of the contractual arrangements which it makes with them. As a consequence of this concentration on the firm as a purchaser of the inputs it uses, economists have tended to neglect the main activity of a firm, running a business. And this has tended to submerge what is to me the key idea in ‘The Nature of the Firm’: the comparison of the costs of coordinating the activities of factors of production.

---

4To be more accurate, differences in production costs are allowed in Williamson’s (1985) approach, but these are, however, primarily in the form of (a) economies of scale that distinguish suppliers from in-house production and (b) differences in technology (specialized vs. generic). Technology is still held constant (Williamson, 1988).
within the firm with the costs of bringing about the same result by market transactions or by means of operations undertaken within some other firm. (Coase, 1988: 38; italics added)

There are two issues here to which I would like to draw attention. First, the main activity of firms is not about buying and selling or about the choice of contractual arrangements but, rather, running a business (of which buying, selling, and choosing contractual arrangements are but a part, albeit an important one). Second, the relevant comparison is not mere costs, but the costs of bringing about the same result, be it similar output at lower cost or superior output at the same level of cost. This underscores that there are differences among firms in the realm of production which need to be recognized in trying to better understand the institutional structure of production (also Langlois and Foss, 1999). The source of firm advantage lies in those activities which it is able to conduct in a superior manner vis-à-vis other firms and which are difficult for other firms to emulate competitively within an acceptable time frame or cost. Such differences would explain why an activity is organized within a particular firm and not obtained through the market (also see Langlois 1988, 1992), the market here implying, as pointed out earlier, organization of the same activity within another firm. Here, Coase’s emphasis on costs is logically equivalent to RB theorists’ arguments on competitive advantage in that such advantage is the logical outcome of a superior cost position (of bringing about the same result), whereas cost is the tool or vehicle to attain this competitive advantage. In a sense, they are two sides of the same coin.

By shifting attention away from just the TC associated with organizing through the market to also incorporate the differences in production costs between firms, the root of which lies in the differing capabilities of firms (also see Langlois, 1992; Langlois and Robertson, 1995), the argument provides a more comprehensive lens into the organization of economic activity. To the extent that most production is carried out within firms and most transactions are firm–firm transactions, Coase reflects upon his original 1937 work in two important ways. First, as elaborated in this section: [T]he dominant factor determining the institutional structure of production will in general no longer be TC but the relative costs of different firms in organizing particular activities. This does not mean that TC will not be important in particular cases nor that they will not be important in determining the form of the contractual arrangements made by firms. What it does mean, if I am right, as I put in my Yale lectures, “to explain the institutional structure of production in the system as a whole”, it is necessary to uncover the reasons why the cost of organizing particular activities differs among firms. (Coase, 1990: 11)

Second, and the focus of the next two sections:

The institutional structure of production comes into being under the influence of forces determining the interrelationships between the costs of transacting and the costs of organizing ... But it is a theoretical scheme that incorporates these interrelationships that I believe will make my approach to The Nature of the Firm operational. (Coase, 1988: 47)

To the extent that this implies that the make-or-buy decision needs to additionally explore the process through which firms produce goods and services (Menard, 1994), it shifts more of the relative burden of explaining the institutional structure of production onto the shoulder of the RB view of the firm.

**TOWARDS A GREATER UNDERSTANDING OF THE INTERRELATIONSHIPS BETWEEN PRODUCTION AND EXCHANGE RELATIONS**

There are numerous reasons why proponents of the two theories—TC and RB—have resisted a dialogue. (See Table 1.) First, in contrast to the theory of the firm that characterizes TC economics, the RB view is the theory of a (particular) firm. That work on the theory of the firm has developed quite separately from the theory of a firm is hardly surprising since they basically address two different questions: (broadly) why firms exist and why they differ (or why there are performance differences across firms), respectively. Their domain of interest is thus correspondingly different: the search for the efficient governance structure and the search for competitive advantage, respectively. Since they are interested in two different aspects of economic activity, exchange and production,
Table 1. Transaction costs and resource-based theories of the firm: a comparison

<table>
<thead>
<tr>
<th>Broad theoretical arena</th>
<th>Transaction cost theory</th>
<th>Resource-based theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary theoretical question</td>
<td>Why do firms exist?</td>
<td>Why do firms differ?</td>
</tr>
<tr>
<td>Primary driver</td>
<td>Search for efficient governance structure</td>
<td>Search for competitive advantage</td>
</tr>
<tr>
<td>Primary domain of interest</td>
<td>Exchange and the transaction</td>
<td>Production and firm resources/capabilities</td>
</tr>
<tr>
<td>Primary focus of analysis</td>
<td>Transaction attributes (e.g., asset specificity)</td>
<td>Resource attributes (e.g., value, stickiness)</td>
</tr>
<tr>
<td>Primary emphasis</td>
<td>(Transaction) Costs</td>
<td>Firm resources, skills, knowledge, routines</td>
</tr>
</tbody>
</table>

respectively, their focus of analysis correspondingly differs—for one, the transaction; for the other, firms’ resources/capabilities. Accordingly, one analyzes transaction attributes (e.g., specificity of assets, measurement difficulty), while the other analyzes resource attributes (e.g., value, imitability, stickiness). Given their quasi-paradigmatic differences and the fact that TC primarily emphasizes cost while RB emphasizes issues to do with skills, knowledge, and routines, it is not altogether unexpected that their respective approaches to understanding firm behavior and economic organization differ and that they have paid inadequate attention to understanding how the two aspects of production and exchange relate to one another.

In spite of their differences, however, there are common issues which link the two. For instance, a firm can be seen as both a collectivity of transactions (Ulrich and Barney, 1984; Winter, 1988) and as a bundle of resources. Governance skills, both within and across firm boundaries, can result in performance differences and competitive advantage (Dyer and Singh, 1998; Adler, Goldofitas, and Levine 1999). When firms transact through exchange, they transact resources (Madhok and Tallman, 1998; Chi, 1994). Here, resource attributes such as tacitness increase the measurement problem and thus impact upon the level of TC (Chi, 1994; Silverman, 1999). If firms are superior to markets for reasons of efficiency, this may well be due not just to TC reductions but to productivity-enhancing factors tied to superior skills and knowledge.

Scholars (Walker and Weber, 1984, 1987; Monteverde and Teece, 1982; Masten, Meehan, and Snyder, 1991) have documented support for both kinds of costs. Walker and Weber (1984) found that both production and transaction costs mattered in the make-or-buy decision, although production costs overshadowed TC. Interestingly, in their later study (Walker and Weber, 1987), TC mattered only when examined in conjunction with supplier competition, suggesting the importance of incorporating contextual variables, an issue I address later. Along similar lines, Masten et al. (1991) found that, although much of the TC work has focused on the costs associated with market exchange, a large part of the make-or-buy decision was explained by variations in internal organization costs, most of these having to do with dissimilarity of skills with respect to the relevant task.

While both sets of costs are undoubtedly important, scholars have nevertheless tended to basically graft TC onto production costs, reflecting the lack of dialogue. This is clearly not sufficient. As a result of their commonalities and interdependencies, the challenge and the opportunity, as indicated by Coase (1988) upon further reflection, lies in uncovering the interrelationships between these two sets of costs and, correspondingly, between hierarchical (production) and market (exchange) relations. Such a theoretical schema would considerably improve our insight into economic organization.

What the RB view enables us to do in this regard, which Coase with his emphasis on costs failed to recognize adequately, is to shift the focus away from cost to firms’ skills, capabilities, and knowledge. In a somewhat neglected paper, Richardson emphasized this very issue:

> It seems to me that we cannot construct an adequate theory of industrial organization and in particular to answer our question about the division of labor between firm and market unless the elements of organization, knowledge, experience, and skill are brought back to the foreground of our vision. (Richardson, 1972: 888)

Along similar lines, Carlsson contends that:
We must get beyond the production function and develop a theory which... consists far less of physical transformation and more of knowledge processing in the form of R&D, engineering, marketing and administration than is commonly assumed in conventional models... It must also make it possible to incorporate management concerns, such as strategy, creativity, and entrepreneurship. (Carlsson, 1992: 5; italics added)

Consistent with this view, the decision with respect to the appropriate governance structure rests not just on costs, but also on productivity benefits tied to skills and knowledge.

The triangular alignment hypothesis: Transaction, governance structure, and resource particulars

Clearly, economic scholars interested in the theory of the firm have tended to restrictively focus on the cost dimension. This has led to an expression of concern about theories such as TC theory which, by being overly preoccupied with economizing on costs and by not adequately accounting for how knowledge is managed, have become rather tenuous explanatory theories about the link between organization and competitive advantage (Liebeskind, 1996). TC theorists argue that the choice of governance arrangements is primarily due to transaction attributes and essentially ignore resource attributes and governance skills, which is the domain of RB theory. RB theorists focus more on the theory of a (particular) firm and on how firm resources and skills can be managed for competitive advantage. The TC theory of why firms exist (i.e., why firms in general would/should organize a particular activity internally) does not fully explain why a particular firm will/should (or would/should not) organize that activity hierarchically within its boundaries. Thus, a more complete theory of why a given firm will (or will not) integrate a particular activity in-house is required.

A truly strategic theory of the firm should address not just the decision with respect to hierarchical governance or market governance, i.e., production or exchange, but also take into account how a firm’s resources and capabilities can best be developed and deployed in the search for competitive advantage. Clearly, the reason why there are variations in organizational form under similar transaction characteristics or, alternatively, why different firms organize similar transactions in different ways is that it is not just transaction particulars that matter, but also firm particulars. In such a case, instead of a more simple bilateral alignment, the central issue becomes a triangular alignment between the characteristics of the governance structure, the characteristics of the transaction, and the characteristics of relevant firm resources (Figure 1). Bringing about

Figure 1. The triangular alignment hypothesis
such an alignment requires greater attention to the interdependence between costs and skills and addresses Coase’s concern regarding the importance of addressing the interdependence between production and exchange relations.

Recent research attests to the importance of addressing the triangular alignment argument in firms’ boundary decisions. For instance, in their study of foreign currency trading, Mosakowski and Zaheer (1999) found the coexistence of two types of populations in the same setting: smaller, single-unit, and geographically isolated firms and larger, multiple-unit firms with global configurations. In contrast to the former, organizational attributes linked to information-processing ability, partly cultivated through greater and more diverse experience, increased the latter’s operational flexibility and made them more adept at foreknowledge exploitation and more flexible in expanding and contracting their boundaries. At the same time, neither strategy was necessarily superior over the other since each had their trade-offs. Compared to the multiple-unit firms, the single-unit firms forewent the coordinating benefits, but were, in turn, able to achieve a greater alignment of incentives and lower agency costs because the resources/capabilities required for the ‘simpler’ strategy were different. What this reflects is that the two different trading strategies corresponded to two different transaction-resource-governance configurations, both of which were in alignment. Clearly, if the small firms governed through mechanisms like the large firms in the absence of similar resource attributes, or vice versa, the strategy would not have been as successful.

In another interesting paper, Miller and Shamsie (1996) made a distinction between discrete and systemic and between property- and knowledge-based resources, and found that control over property-based resources resulted in superior performance during periods of stability while control over knowledge-based resources resulted in superior performance during periods of change. The type of resources/capabilities relevant to a particular activity therefore matters in an important way. For instance, in the systemic case, rather than ownership of a particular resource, the key issue for competitive advantage could well be the enlargement of the range and comprehensive-ness of the system. This has different implications for governance than in the discrete case. Following this line of argument, a firm treating an investment, say a distributorship, as a discrete property may approach the transaction differently than another firm which approaches it as part of a larger distribution system, one which enables it to attain a more economic allocation of overheads throughout the system, share administrative and other skills systemwide, and better exploit other attributes such as brand image or reputation.

What the above suggests is that the identity and the strategy of the particular firm also influence how its resources interact with the transaction and how it chooses to govern it. Since different firms have different capabilities and different strategies (hopefully) in line with their capabilities, it can be expected that they will organize their activities differently. While transaction particulars may remain unchanged, for instance, a transaction characterized by high asset specificity remains independent of who organizes the transaction, the value of a particular resource attribute depends on the particular firm within which it is housed and how the firm applies this attribute to the transaction. This is a function of such idiosyncratic characteristics as its history, routines, configuration, other resources, etc., which provide the surrounding support structure.

In seeking triangulation, both TC and RB theory need to pay more concerted attention to the context within which the activity occurs (Oxley, 1999; Priem and Butler, 2000). The value of a particular resource is dynamic and changes over time. Therefore, for instance, where a firm’s competitive advantage is rooted in control over a property-based resource, say a patent or technological standard, a shift in the environment to more knowledge-based resources, such as design skills, integrative skills, and the like, can erode the resource value. The result can be a triangular misalignment, reversing any prior alignment and calling for a new alignment. IBM’s conversion from a reliance on in-house production in the 1980s to that of alliances in the 1990s as the computer industry modularized is an example of this.

The above illustrations mainly had to do with different configurations within the firm. One can also apply the argument more generally across firm boundaries. For instance, one finds more alliances in high-velocity environments (Hagedoorn, 1993), one reason being that past resource endowments may be less valuable and past routines result
in inertia. Speed and flexibility are particularly important capabilities for competitive advantage in such environments. It can be argued that one reason why a specialized investment does not lead to vertical integration in such environments is that the triumvirate of transaction attribute (asset specificity), the requisite resource/capability attribute (speed/flexibility), and the governance attribute (low incentive intensity) would be misaligned. The speed/flexibility argument also explains why such environments are characterized by a greater prevalence of smaller firms or why large firms spin off some of their more dynamic businesses or manage them separately.

In contrast, in a situation of technological complexity and uncertainty and competitive intensity, alliances may occur for a couple of reasons. Firms may recognize that, rather than risk premature obsolescence of investments, they can access others’ specialized investments. Additionally, not only can firms complement their resources with those of other specialists, but, by also structuring the more microlevel governance mechanisms appropriately, the collaboration can enable the transfer of skills and knowledge, which in turn may add to a firm’s capabilities with respect to speed/flexibility and thus augment its competitiveness. In this situation, the lower criticality of the transaction attribute (specialized investment) due to the risk of obsolescence (Balakrishnan and Wernerfelt, 1986), the attributes of the governance form (greater incentive intensity and more diverse knowledge stock), and the requisite resource/capability results in greater alignment. At a more particular level, different firms have different propensities to cooperate and also differ in how they cooperate (see next section).

Let us take a different example. Walker and Weber’s (1984) hypothesis that buyer experience with production would reduce the production cost advantage of the supplier and result in a make decision was not supported. There was also no support for the alternate hypothesis that buyer experience would reduce information asymmetries and thus lower TC and result in a buy decision. The authors speculated that the simplicity of the component might explain the lack of significance. Ostensibly, where the component and the associated skills/capabilities are more complex, the steeper learning curve would provide a greater production advantage to the supplier over an inexperienced buyer. In the simple case, the lack of complexity makes the skills/capabilities relatively generic or easily available (i.e., not attached to a specific firm) and the lack of experience does not occasion much of a disadvantage over a supplier firm. At the same time, there is not much of a transaction disadvantage with respect to information asymmetries, compared to the more complex case, since the amount and the complexity of the relevant information is relatively minimal. As a result, the simplicity of the resources involved in this case results in an ambivalent impact with respect to both production and transaction criteria and renders the make-or-buy decision somewhat indeterminate in that alternate governance structures could be equally viable (aligned) with resource and transaction attributes without any particular advantage or disadvantage.

To sum up, firm strategy and resource particulars have an important role in the decision about firm boundaries. Even Williamson acknowledges the importance of firm particulars in his recent paper (Williamson, 1999). Making a significant departure from his earlier near-exclusive emphasis on transaction characteristics (Williamson, 1991b), he concedes that there is merit in shifting the emphasis away from the best generic institutional form for organizing a particular transaction to the best way for a specific firm—with its history, routines, resource endowments, local institutional context, etc.—to organize this transaction. This, in essence, subscribes to the triangular alignment hypothesis above.

With respect to firm boundaries, not only are the capabilities required often difficult to fully attain internally, especially in dynamic environments, but they are also difficult to transact well at arm’s length. In the following section, I argue that where two sets of resources need to be commingled in an uncertain way, as is especially the case in situations of innovation, costs become more variable. Here, it is not just production but also governance skills that become important in order to align costs and competencies with the governance structure.

**THE CASE OF INTERFIRM COLLABORATION**

The shift in focus from just cost to also incorporate skills and knowledge enables us to examine the interrelationships between production and exchange with a fresh lens. When one examines
the activities of firms, it is evident that there are two kinds of skills: production skills and governance skills. Firms can tap into their own production skills, those of another, or a combination of the two. I examine this issue in more detail in this section, in the context of collaborative relations between firms. These have become so pervasive in recent years that they are arguably reshaping the very institutional structure of production (Teece, 1992).

Generally speaking, collaborations enable firms to attain some outcome which they are unable to attain on their own, at least not within the same time frame or cost level (Madhok and Tallman, 1998). In line with their different orientations, the lens that the two theories apply to interfirm collaboration is basically different. From the RB perspective, interfirm collaboration occurs because it enables a firm to access complementary resources and thus overcome RB constraints (and hopefully increase productivity) which are needed to sustain its growth, whereas, from the TC perspective, interfirm collaboration occurs only when it minimizes the cost of governing that activity (Combs and Ketchen, 1999).

Interfirm collaboration in its various forms is typically characterized by a blurring of boundaries between production and exchange, especially in situations in which innovative activity is involved, as firms increasingly eschew pure market relations and engage in closer relations with a limited number of economic actors. In such a context, firms are not only transacting resources through arm’s-length relations, but are also playing a joint and participatory role in the production of such resources. The more central issue here is not internal production or external exchange, but rather the extent of ‘arm’s-lengthness’ of the relations, which, in a sense, reflects the degree of separation between production and exchange.

The example of the NUMMI joint venture between General Motors and Toyota, elaborated on at length by Adler and his colleagues (e.g., Adler, 1993; Adler and Borys, 1996; Adler et al., 1999), exemplifies the central point and serves as a useful vehicle to illustrate the key arguments further. Adler highlighted the following points:

1. The U.S. automakers saw the role of the supplier as one of fulfilling the terms of the purchase contract and meeting the product specifications provided by the company. In contrast, NUMMI expected continuous improvement and innovation on the part of their suppliers and worked more closely with them to develop their capabilities.
2. Unlike the U.S. firms which, being distrustful of their supplier’s competence, closely inspected incoming parts, NUMMI certified their ability.
3. Instead of departures from procedure being treated as a threat to be circumvented, NUMMI treated them as opportunities for learning.
4. U.S. automakers maintained flexibility in their dealings with suppliers through short-term contracts with many suppliers. While this required them to do the design work in-house and then call for competitive bidding, it cut them off from suppliers’ knowledge (and investment). In Toyota’s case, it was the reverse: just a couple of suppliers per part and emphasis on joint design and development. This entailed a redistribution of tasks across firm boundaries.

The net result was that while the Big Three followed a cost-minimization strategy, grafting transaction costs onto production costs, and pursued a narrow kind of flexibility, in terms of ease of switching between suppliers, this strategy precluded the design and efficiency improvements that could have been attained through suppliers’ involvement. Thus, there was a kind of trade-off between flexibility and efficiency. This trade-off was minimized in the case of NUMMI since, rather than treating production and exchange as separate endeavors, it envisioned the relationship as both an exchange (transacting) as well as a production (productive) endeavor and invested in it accordingly. As a result, NUMMI was able to push the efficiency–flexibility frontier outward rather than just move along the curve.

The NUMMI strategy tied them more closely to their suppliers. In fact, the process of conveying mutual needs, competencies, and expectations and making the necessary mutual adaptations results in the investment becoming one of a progressively specialized nature, characterized by the potential for appropriative behavior. Yet, at the same time, without tying together the respective firms’ relevant assets to the specialized application, the benefits of doing so may end up being sacrificed. Specialized investments are therefore a double-edged sword, the net effect from an efficiency perspective being somewhat ambiguous (Poppo and Zenger, 1998).
In contrast to the archetypical TC reasoning, however, such specialized investments need not drive a firm to integrate vertically. In the above example, close relationships and intense interaction can not only lower transaction and/or production costs, but also have productivity benefits. For example, working together from the outset on joint design and development would result in fewer errors, improved quality, more rapid dissemination and absorption of information and skills, reduced development time, and, consequently, lower production costs, more so than the individual firm would have the capacity to do on its own (see also Clark and Fujimoto, 1991). Moreover, a combination of (a) having fewer suppliers, (b) having larger volumes and repeat transactions with these suppliers and (c) sharing more information with them, (d) paying attention to the fostering of trust, and (e) making relationship specific investments that bind one another enabled NUMMI to maintain their relationships long term in a manner enabling lower transaction costs than the American firms, in spite of higher asset specificity (also Dyer, 1997).

Thus, the trade-off between production and governance efficiency (Poppo and Zenger, 1998) may be somewhat spurious, in which case the governance decision from a purely cost perspective becomes somewhat indeterminate and, additionally, depends on the production and relational skills of the participants. The focus on cost alone is limiting since it ignores the opportunities that can be attained and deployed through intensive relationships (Lorenzoni and Lipparini, 1999). As Loasby puts it:

What is at stake in most of these relationships is more than economizing of the transaction costs ... such intimate connections are forms of organization which aid the growth of knowledge by providing effective frameworks—or effective links between distinctive frameworks—within which conjectures can be formulated, criticized, tested, amended, and superseded. (Loasby, 1994: 261)

From this perspective, the focal actors become both participants in a productive endeavor as well as transactors in exchange. This acknowledges the interdependence between costs and skills.

The above argument has a number of important implications, all related to one another in one way or another. First, from a ‘bare bones’ TC perspective, the choice regarding mode of production is fairly simple—either internalize or use less specialized investment. Yet, such an approach deters from a more detailed and fine-grained analysis of differential firm skills and capabilities. The question with respect to internalization or to the nature of collaborative relations, i.e., more vs. less arm’s length, depends to some extent on a firm’s judgement of the value of the relevant productive resources of the two parties, independently and interdependently, and of the potential impact of the exchange on the firm’s existing resource/capability pool and resource productivity.

Second, and related to the first, where there is the possibility that the production skills of one firm are further enhanced by the production skills of another, through learning, the extent of this depends on the governance skills (i.e., the skills involved in structuring and managing the exchange relationship) on the part of one or both parties. These production and exchange skills influence both the opportunities as well as the ability to learn from one another, which in turn could result in both lowering costs and/or enhancing the productivity of the internal resources. This line of argument draws attention to the contention that, even though organization form choices, whether internal or external, may (or may not) be explained by TC theory, yet ‘it may be more informative to focus on the additional value that can be created within an administrative framework which facilitates interaction and thereby creates opportunities’ (Loasby, 1996: 47).

Third, and related to the second, the behavior and performance of firms engaging in interfirm collaboration for RB reasons of enhancing competitive advantage may well be different from those doing so for TC reasons where the dominant orientation is governance efficiency and the control of organized effort (Madhok, 2000a). In the former case, rather than being an independent entity with whom one transacts, the supplier becomes a resource and a strategic asset, complementing and enabling firms to consolidate in-house competencies. It is important here to distinguish between the entrepreneurial aspect of the exchange and the managerial one, with the former endeavor being concerned more with creation and the latter with execution. The latter may enable cost improvements, but would be unable to stimulate the creation of new knowledge (Adler, 2001).

Fourth, the propensity for other-oriented behavior motivated by joint interest coexists with more egoistic behavior concerned with self-interest. The
relative importance of the two kinds of behavior is variable and depends on the circumstances, since the parties’ preferences are in actuality endogenous rather than exogenous to the social process (Adler, 2001). For instance, rather than resort to contractual mechanisms in a situation of weak compliance, a firm can attempt to better identify and appreciate the factors that might be hindering better compliance (e.g., genuine differences in comprehending each other’s needs and competencies) so that friction may be reduced. Even if exchange conditions were exogenous, endogeneity of preferences would still determine how a firm responds to such conditions in a transacting (exchange) relationship. This endogeneity of preferences also influences the nature and extent of specialized investments that a firm is willing to commit, which in turn has an implication for the productivity of those investments.

Fifth, in TC theory, the key issue of a ‘thin’ market, resulting in lack of competition among suppliers (Walker and Weber, 1984), has resulted in the neglect of ‘the heterogeneity of potential partner capabilities, ex ante’ (Oxley, 1999: 23). This is an important point since a collaboration with a particular partner may make sense, while it may not make sense with another, even though transaction characteristics remain the same, simply because the rents attached to a particular resource combination may be particularly valuable. Moreover, it is not just an issue of collaboration, but how one collaborates which will bring about the transaction–resource–governance alignment.

Sixth, the driving issue is not the TC but the net resource value of a transaction. While the act of binding two sets of resources more closely together can be costly and also of a relationship-specific nature, much more so than if it was just a control effort, the value of the transaction is derived from the net economic surplus through jointly tying together production and exchange relations, which could more than offset the associated costs (Argyres and Liebeskind, 1999; Madhok, 2000b, 2002; Oxley, 1999) as a result of the interdependence between costs and skills. In this process, some current expenditures that may appear discretionary from a static perspective are still dynamically essential and are akin to a strategic use of costs. In other words, firms could well increase the overall cost associated with transacting, including knowledge transfer costs, in the initial stages, without an overall cost penalty, since smoother coordinating interfaces and more mutually aligned skills, knowledge, and expectations between the parties could lower production and/or transaction costs in the longer term (Langlois, 1992). What this suggests is that rents due to superior performance as a result of enhanced productivity of resources can subsidize what may otherwise be considered transaction inefficiencies.

The NUMMI example demonstrates that the interdependence between production and exchange occurs not just in the area of costs, but also in terms of knowledge and learning outcomes. In general, when skills and knowledge are involved, there is a coevolution between the competencies of firms and those of their partners (Lorenzoni and Lipparini, 1999). It is important to recognize this coevolution.

Linking back to the previous section regarding a particular firm, interfirm collaboration occasions changes in a firm’s productive opportunity set. If the firm possesses the appropriate governance skills, it is no longer limited to just its production technology, but has a broader set of production technologies to choose from. The extent to which one firm has a greater range of production technologies to choose from than another firm because of its relational skills and/or because it is able to better benefit from such choices can become a source of performance differences across firms, especially when cooperation can put the firm onto a different dynamic path than a go-it-alone or more arm’s-length strategy. Such competitive advantage is based not purely on efficient governance, but also on the impact of governance skills on the productivity of firm resources (Dyer, 1997; Adler et al., 1999; Nishiguchi, 1994; Helper and Sako, 1994). Thus, skillful governance is a type of capability in and of itself.

In regard to the triangulation alignment argument presented in the previous section, it must be recognized that the transaction costs incurred in the exchange of productive resources are not independent of the nature of resources to be transacted and, similarly, the returns realized from these resources are not independent of the transaction costs incurred (Madhok and Tallman, 1998). Therefore, in a situation of identical transaction characteristics, which was the case in the NUMMI example, the U.S. auto firms, in line with their strategy, aimed for low costs, a low alignment of technologies to choose from than another firm because of its relational skills and/or because it is able to better benefit from such choices can become a source of performance differences across firms, especially when cooperation can put the firm onto a different dynamic path than a go-it-alone or more arm’s-length strategy. Such competitive advantage is based not purely on efficient governance, but also on the impact of governance skills on the productivity of firm resources (Dyer, 1997; Adler et al., 1999; Nishiguchi, 1994; Helper and Sako, 1994). Thus, skillful governance is a type of capability in and of itself.

In regard to the triangulation alignment argument presented in the previous section, it must be recognized that the transaction costs incurred in the exchange of productive resources are not independent of the nature of resources to be transacted and, similarly, the returns realized from these resources are not independent of the transaction costs incurred (Madhok and Tallman, 1998). Therefore, in a situation of identical transaction characteristics, which was the case in the NUMMI example, the U.S. auto firms, in line with their strategy, aimed for low costs, a low alignment of technologies to choose from than another firm because of its relational skills and/or because it is able to better benefit from such choices can become a source of performance differences across firms, especially when cooperation can put the firm onto a different dynamic path than a go-it-alone or more arm’s-length strategy. Such competitive advantage is based not purely on efficient governance, but also on the impact of governance skills on the productivity of firm resources (Dyer, 1997; Adler et al., 1999; Nishiguchi, 1994; Helper and Sako, 1994). Thus, skillful governance is a type of capability in and of itself.
NUMMI pursued a ‘higher-order’ strategy of value creation through joint entrepreneurship. The latter required a greater alignment of skills between the parties, even though it entailed higher costs to attain these skills, and governance mechanisms to support this process. Both strategies coexisted even though it could lead ultimately to differences in performance. However, one could argue that, if NUMMI aimed to attain its strategy through the mechanisms used by U.S. producers, or vice versa, it would have failed due to a misalignment between the transactional, resource, and governance attributes.

To connect the argument back to Coase, one can on the one hand examine production and transaction costs separately, in which case, production costs are grafted on to TC. On the other hand, the net value argument approaches costs not just in terms of the sum of production and transaction costs, but also incorporates their interaction, where the magnitude and nature of the costs incurred interact to yield overall economic value or surplus (or deficit) through the relationship between cost and skills. This underlines the central point that, to comprehend the institutional structure of production, one needs to more closely scrutinize and understand the interdependence between transaction and production relations.

CONCLUDING REMARKS

In this paper, I explained why Coase’s contribution is even more central to strategy scholars than previously recognized. The discussion in this paper probed some of the interrelationships between transaction and production relations which, as can be seen, can be quite complex in nature. Clearly, strategy has a major role in decisions about firm boundaries, in terms of choice as well as nature. In this regard, Rumelt, Schendel, and Teece’s (1991: 19) contention that ‘strategic management is about coordination and resource allocation inside the firm’ can be extended further. I would contend that, as a result of the interdependence of production and exchange relations, strategic management is about coordination and resource allocation both within and across firm boundaries.

In his 1991 paper, Williamson stated with respect to business strategy that ‘What is missing in business strategy, but is desperately needed, is a core theory’ (Williamson, 1991b: 90). He elevated (transaction cost) economizing over strategy considerations and argued that ‘the microanalytic comparative institutional, economizing orientation of transaction cost economics deals with many of the key issues with which business strategy is or should be concerned’ (Williamson, 1991b: 90). And yet, on closer examination, he missed the central issue. While TC economics has undoubtedly made important contributions to strategic management theory, particularly in the realm of economic organization, it is nevertheless only a partial solution since it provides, at best, a tenuous link with competitive advantage, arguably the key issue of concern for strategy. Clearly, the theory of the firm and the theory of performance differences between firms are not unrelated. This paper elevates strategy considerations to a much more central position in understanding the institutional organization of production. The triangular alignment hypothesis offers scope for a more comprehensive theoretical solution.

In recent years, a number of authors have argued that firms are not just efficient governance structures, but also institutions for learning (Teece, 1990; Kogut and Zander, 1992; Ghoshal and Moran, 1996; Madhok, 1996, 1997). In this light, governance structures—and associated mechanisms—serve to not only align transaction and governance characteristics, but also serve as a vehicle to manage skills and knowledge. This underlines the importance of the ‘third leg of the stool.’ To provide a broader framework to understand both governance and sources of competitive advantage, ‘TCE should be coupled with RBV and KBT (knowledge-based theory) assumptions that integration of competencies and combinations of knowledge across a firm’s boundaries is an important goal of managerial action’ (Lorenzoni and Lipparini, 1999: 332).

While recent studies which look at both TC and RB variables (Poppo and Zenger, 1998; Silverman, 1999; Combs and Ketchen, 1999; Argyres, 1996) find that both matter, they have tended to address the interdependence between the two sets of concerns in only a limited manner, if at all. For instance, they do not address how capabilities can co-evolve through governance skills. Without closely and more explicitly examining the role of skills, the relationship between costs and skills, and how the production skills of firms can co-evolve with governance skills, the gains from
studies attempting to bridge the two theories will remain limited.

In this paper, I have attempted to show the importance of shifting the focus from a bilateral alignment to a triangular alignment between the triumvirate of transaction, governance structure, and resource characteristics. More concerted effort needs to be made in this direction. For instance, ‘Which firms, with what attributes, are more (or less) competent in developing or deploying what kind of resources/skills/knowledge to further their productive capacity, while simultaneously protecting its value, and in what way or in which kinds of contexts? Or, which firms, with what attributes, are able to learn what kinds of knowledge and in which kinds of contexts in a way that enhances their productive capacity?’ This would be a promising area for future research.

Besides the choice regarding institutional form, the paper also showed how, even within institutional choices, the microstructure of the institutional arrangements can both enhance the productivity of resources or reflect TC concerns or both. This is especially important when a particular phenomenon, such as specialized investments, does ‘double duty,’ creating opportunities as well as constraints. The same is true for tacitness. Under what circumstances should/will a firm’s production and/or transacting ‘technology’ tilt towards one ‘duty’ (opportunity maximization/value creation) or the other (opportunism minimization/value appropriation)? How is this calibrated? In what way does this depend on the particular combination of resource, transaction, and governance characteristics? Is there also an evolutionary or co-evolutionary aspect to this calibrating process and, if so, how does this function?

On a different note, TC theory approaches hierarchical and nonhierarchical forms such as collaborations as substitutes. Yet, they could also be approached as complements, especially in situations involving learning (Cohen and Levinthal, 1990; Madhok and Osegowitsch, 2000). How does one reconcile this? Under what circumstances are they better viewed as substitutes and under what circumstances as complements? Perhaps addressing the interrelationship between costs and skills and addressing the triangular alignment argument would provide some answers.

In general, what is the nature of the relationship between costs and skills and how do production and transaction costs go together? How do they leverage one another? How can they be mutually beneficial? When and under what circumstances? What are the trade-offs between production costs and transaction costs; between the short term and the long term; between the static and the dynamic aspects; between current expenditures and future revenues; between safeguarding against opportunism and foregoing of opportunities; etc.? How are they made? When and under what circumstances? These are important questions which beg closer attention and need to be addressed within a more comprehensive theoretical framework in order to more fully understand and explain the institutional structure of production. Clearly, strategic management scholars have a pivotal role to play in this regard.

ACKNOWLEDGEMENTS

I would like to thank Steve Tallman, Bill Hesterly, Jeff Dyer, and three anonymous reviewers for their feedback. Their insightful comments helped tremendously in sharpening the arguments presented in the paper.

REFERENCES


