Capabilities, Transaction Costs, and Firm Boundaries: A Dynamic Perspective and Integration

Nicholas Argyres*
School of Management
Boston University
595 Commonwealth Ave.
Boston, MA 02215
Phone: (617) 353-4152
Fax: (617) 353-5003
nargyres@bu.edu

and

Todd Zenger
Olin Business School
Washington University
One Brookings Drive
St. Louis, MO 63130-4899
Phone: (314) 935-6399
Fax: (314) 935-6359
zenger@wustl.edu

*We express appreciation to Dan Elfenbein, Rich Makadok, Jackson Nickerson, Joanne Oxley, and Lamar Pierce for helpful comments. Previous versions of the paper were presented at the Academy of Management Meetings in 2005 and 2007.
Abstract

A large literature has developed in recent years that attempts to compare transaction cost and capabilities explanations of firms’ vertical boundaries. Much of this literature has treated comparative capabilities (buyers’ vs. potential suppliers’) as determinants that are independent of transaction costs, based on the idea that capabilities theories of the firm are distinct from the transaction cost theory of the firm. We argue that this approach is mistaken. We contend that capabilities and transaction cost determinants interact with each other dynamically, and that the two theories of the firm cannot be conceptually distinguished. We then seek to articulate an integrated perspective that incorporates both capabilities and transaction cost logic. Our argument carries implications for theories of the firm, and for empirical research aimed at testing those theories.
Introduction

In recent years an active theoretical and empirical debate has emerged around the topic of the boundaries of the firm. While transaction cost economics (Williamson, 1975; 1985; Klein, Crawford, & Alchian, 1978) has long dominated this debate, particularly in the empirical literature, scholars have more recently advanced competing arguments based on capability and resource-based logic (e.g., Argyres 1996; Conner & Prahalad 1996; Foss 1996a, 1996b; Poppo & Zenger 1998; Madhok 2002; Leiblein & Miller 2003; Jacobides & Hitt 2005). In particular, scholars have highlighted a distinctive role that firms’ comparative capabilities play in defining boundaries of the firm (Walker & Weber 1984; Argyres 1996; Barney 1999; Jacobides & Hitt 2005; Jacobides & Winter 2005; Madhok 2002). Most of this work positions capability explanations as independent of, and often competing with, more traditional transaction cost explanations for the boundaries of the firm. On the surface, these two theories do seem to propose starkly different explanations for boundary choice (e.g., Coombs & Ketchen 1999; Leiblein & Miller 2003; David & Han 2004; Jacobides & Hitt 2005; Jacobides & Winter 2005).

Transaction cost logic argues that activities are internalized when the cost of governing the activity through the market exceeds the cost of governing it with the internal hierarchy of the firm. By contrast, capabilities logic explains the choice to internalize (or the persistence of this choice) as a reflection of superior capability to perform the activity within the firm relative to the capabilities of external providers.¹

¹ Jacobides and Hitt (2005), for example, state that, “Firms with greater productive capabilities in a stage of production will tend to perform this activity internally, and contract with another firm through the ‘market’ where they are deficient. When skills or knowledge create an advantage for only one segment of a value chain, firms will tend to specialize; if these skills apply across multiple value chain segments, there is little basis for specialization at the firm or industry level” (p. 1210).
Articulated in this manner, these arguments appear quite distinct. However, as we will argue, a more careful look reveals considerable difficulty in conceptually distinguishing them.\textsuperscript{2} As evidence of this difficulty, scholars from both perspectives seem to easily interpret the same empirical results as supportive of their preferred theory (David & Han 2004; Carter & Hodgson 2006). Thus, a recent review of the empirical literature in transaction cost economics concludes that the majority of empirical make-or-buy studies testing TCE hypotheses can be “reinterpreted as … consistent with a competence or resource-based perspective (Carter & Hodgson 2006: p. 473.).” Our contention is that this conclusion is not surprising precisely because these theories’ predictions about the basic boundary choice are very close to each other conceptually. Thus, our agenda in this paper is to first explore the theoretical relationships between the two theories of the firm, particularly highlighting the incompleteness of arguments that rely solely on comparative capability, and then articulating an approach to boundary choice that is consistent with the concerns of both perspectives. We thereby respond to recent calls for better integration of existing theories of the firm (e.g., Mahoney & McGahan 2007).

We begin by noting that transaction cost theory, as developed by Williamson (1975, 1985) and Klein, Crawford and Alchian (1978), speaks directly to the question of how firms efficiently develop capability. This point has not been fully appreciated in the literature, in part because of differences in language. For example, scholars in the TCE tradition including Coase (1937), Williamson (1975, 1985), and Klein, Crawford, &

\textsuperscript{2}The tendency to pose transaction cost and capabilities explanations for governance choices as competitive also appears in the literature on strategic alliances (for a critique, see Oxley 2002). For example, alliance structures are sometimes said to be chosen over arm’s length contracts not for transaction cost reasons, but because they facilitate the development of interfirm routines for joint learning and knowledge transfer better than arm’s length agreements (Zajac & Olsen 1993; Sobrero & Roberts 2001).
Alchian (1978), have generally not used the term “capabilities” in their work. Nevertheless, TCE’s treatment of the canonical make-or-buy decision is easily applied—and indeed, was arguably meant to be applied—to this fundamental question of how to efficiently govern capability development. While firms constantly seek new capability, they commonly integrate as a means to efficiently form or protect it.

Next, we argue that a primary source of confusion and a clear impediment to integration has been the failure to adequately address time and temporal sequencing. We contend that to fully understand the inseparable nature of capabilities and transaction cost theories of firm boundaries, we must examine the boundary choice in a dynamic context—exploring both the origin of a boundary decision and its persistence. Unfortunately, empirical explorations of the two theories have been mostly cross-sectional in nature, and these dynamic issues have therefore been ignored. We illustrate our points with an example from the Ford Motor Company in its early years. Following this, we ask whether an explanation for a boundary choice that is based entirely, or in large part, on capabilities considerations can provide sufficient explanation for a given make-or-buy decision. We argue that the answer is “no”, because any explanation based on capabilities alone (or on capabilities as the determining consideration) must consider why a firm doesn’t simply acquire any comparative capability it lacks, or divest itself of any comparative capability it possesses, for example, in the market for corporate control. If a firm cannot efficiently acquire capability or divest it, then we contend there are transaction cost reasons lurking as the explanation, implying that any capabilities explanation cannot stand alone.
Our next step is to argue that capabilities and transaction cost considerations, because they are so intertwined, should be integrated as part of a single theory of firm boundaries. We suggest that developing such a theory requires setting aside what we view as a logically unsustainable debate around the distinctive roles of capability and transaction costs in explaining firm boundaries, and focusing instead on clarifying the theoretical relationships between asset or activity complementarity, resource co-specialization, and governance choice. We conclude the paper by presenting our own effort at clarifying these relationships.

**Capability Development, Specific Assets and Integration**

The resource-based view of the firm has primarily functioned as a theory of firm performance, arguing that the stock of unique, valuable, and difficult-to-imitate resources possessed by the firm determines its competitive advantage (Wernerfelt 1984; Barney 1986, 1991; Peteraf 1993). However, more recently resource-based logic has been extended beyond explanations of firm performance to explanations of the configuration and boundaries of the firm. The proposed argument is quite simple: firms internally govern comparative capability and outsource access to capabilities where the firm is comparatively incompetent. Thus, Barney (1999) argues that “some firms are simply better than others at doing some things” and therefore the “capabilities possessed by a firm and by its potential partners … should have a significant impact on boundary decisions” (p.138). Similarly, Jacobides and Winter (2005) argue that “to understand when firms are integrated and when they are not,” we must “look at the distribution of productive capabilities” (p. 398).
Empirical research seemingly corroborates this comparative capabilities logic. For example, Argyres (1996) found several examples from a manufacturing firm in which relative firm capabilities appeared to be important drivers of vertical scope. Other scholars have empirically linked the presence of skill sets (Poppo & Zenger, 1998) and specialized experience or expertise (Leiblein & Miller 2003; Hoetker 2005) to vertical integration decisions. Jacobides and Hitt (2005), in examining the evolution of firm boundaries in the mortgage banking industry, similarly found that firms with greater productive capabilities in one stage of production performed this stage internally, and contracted with an outside firm for those stages for which their capabilities were lacking. Thus, the fundamental logic that emerges from the application of resource-based logic in this context is the simple concept that firms determine make or buy decisions through a process of comparative capability assessment (Jacobides & Hitt 2005). As these comparative capabilities shift, so do firm boundaries.

We contend that this comparative capability logic, including that articulated in our own work, is incomplete. In particular, we argue that the comparative capabilities logic faces two theoretical shortcomings in providing an explanation of firm boundary choices that is distinct from transaction cost economics. First, it does not adequately address the role that past governance decisions played in generating capability differences in the present. Second, it does not explain the persistence of boundary choices for developing a given capability regardless of the governance arrangements chosen at its origin. These two shortcomings have their origins, we suggest, in an overly narrow reading of transaction cost theory, which, we argue, is actually vitally concerned with capability development. We thus begin by offering a broader interpretation of transaction cost
theory that we argue is quite consistent with the original articulations of the theory. We then describe the two theoretical problems, and offer a more integrated treatment of boundary choices that aims to address them.

Transaction Cost Economics and Capability Development

Transaction cost economics (TCE) addresses the make or buy decision using a distinctly different theoretical language than the logic of comparative capability. TCE highlights the comparative efficiency of governance forms. The theory argues that hierarchies possess distinct advantages over markets in overcoming fundamental hazards in exchange, specifically hazards which arise when desired exchange requires exchange-specific investments by buyers and/or sellers. While this focus on comparing exchange hazards rather than comparing capability suggests a wide divergence in logic, a more careful look reveals the considerable difficulty in distinguishing them conceptually. We contend that TCE, even as it was originally formulated, is appropriately viewed as explaining the efficient governance of “unique” or “firm-specific” capability development. Thus, the inability to effectively manage the process of unique capability development through the market prompts managers to integrate. Integration in turn promotes the co-specialized or firm-specific investments necessary to generate comparatively superior capability. TCE further argues that integration is ultimately limited by the firm’s inability to provide market-like incentives to motivate the development of broadly applicable (non firm-specific) capability. As we argue below, the

---

3 Note that our critique in this paper is not focused on the knowledge-based theory of the firm literature (Kogut & Zander 1992; Conner & Prahalad 1996; Madhok 2002), which explains the boundary of the firm using logic that goes beyond the simple comparative capability argument. The argument in this literature focuses largely on the advantages that integration possesses in facilitating the learning required in capability formation. See below for further discussion of this argument.
exchange conditions that elevate transaction costs in markets and encourage integration are precisely those that enable capability development.

We define a firm’s comparative capability as its capacity to deliver a product or service at lower cost and/or with superior quality or features than alternative providers. TCE views investments in an exchange as the origin of such firm-specific capability. Exchange-specific investments made either within the firm or by external exchange partners generate unique capability or, as Williamson (1975: p. 28) describes, “non-trivial cost advantages”. These firm-specific investments, however, also create hazardous exchange when the development of capability is attempted across a market exchange. Thus, in the presence of such investments that aim to develop capability, exchange partners are “inclined to expend considerable resources bargaining over the price at which the exchange is to take place” (Williamson 1975, p. 28). Essentially, both parties will bargain over the ownership of this new-found “capability” and the stream of rents that it generates. Thus, when capability emerges in an exchange, it transforms what began as an exchange between a buyer and one of many homogenous sellers into an exchange between a buyer and a highly capable seller. This emerging capability to use Williamson’s language, “…transforms a large-numbers exchange condition …at the outset …into a small-numbers exchange relation on account of idiosyncratic experience associated with contract execution” (p. 29). Moreover, Williamson argues that the reason “outsiders are not on parity with insiders [in term of costs] is usually because outsiders lack firm-specific, task-specific, or transaction specific experience. Such

---

4 Lippman and Rumelt (2003) correctly note that “recontracting problems [can] arise, absent investment, whenever any agent’s ex ante outside value differs from the ex post outside value” (p. 1079). Thus, rents can emerge simply through exchange and joint production in ways that are completely unanticipated ex ante.
experience is a valuable resource and can be used in strategic ways by those who, by
being awarded initial contracts, have acquired it” (p. 31). Thus, consistent with
Williamson, integration is the means by which capability is efficiently developed and
protected from appropriation. In this sense, firm governance is adopted to generate
capability that would not be formed were market governance adopted.

The relationships among capability development, firm-specific assets, and
hierarchical governance are clarified further in Klein, et al.’s (1978) articulation of the
transaction cost argument. These authors focus on the development of what they term
appropriable quasi rents—rents generated in an exchange that result from co-specialized
investments. Appropriable quasi-rents are defined as the value of using the firm’s assets
in the current exchange compared to the value of using these same assets in an alternative
exchange. Thus, Klein et al.’s (1978) fundamental prediction is that “as assets become
more specific and more appropriable quasi-rents are created (and therefore the possible
gains from opportunistic behavior increases), the costs of contracting will generally
increase more than the costs of vertical integration” (p. 298).

We contend that these concepts of specific asset investments and appropriable
quasi rents are very closely related to, if not synonymous with, the resource-based
concepts of firm specific capability and their associated rents. In theory, a firm-specific

---

5 Zott and Amit (2006) articulate the opposite argument; that there is indeed a distinction between the
concepts of firm-specificity and asset specificity, and therefore between transaction cost and capabilities-
based theories of firm boundaries. Wal-Mart provides the key example they use to illustrate their argument.
Zott and Amit (2006) argue that Wal-Mart’s logistics capability is an example of high firm specificity, but
low asset specificity, and that because the capability is internalized, the example is explained by
capabilities but not transaction cost explanations. Their argument, however, only considers asset specificity
in relation to exchanges between Wal-Mart and its suppliers of consumer goods. They do not consider the
asset specificity involved in Wal-Mart’s exchanges with suppliers of other inputs, such as logistics
software, equipment, labor, store management, real estate, etc. We contend that there is considerable asset
specificity within the wide range of activities that generate not only Wal-Mart’s logistics capability, but a
capability can reside either within the boundaries of a focal firm or within the boundaries of a supplier. Thus, a firm specific capability exists when either a focal firm itself, or its supplier, enjoys an advantage (cost, quality, or otherwise) in providing the output of a particular activity to the focal firm. The magnitude of this firm specific capability is directly measurable as the level of the appropriable quasi-rents from that activity—the value of using these assets to service the focal firm relative to the value of using them to service the next highest valuing buyer. Specific asset investments are made in an exchange precisely to promote the development of capability that produces these quasi-rents. Therefore, the magnitude of capability and its specificity to the focal firm is closely related to, if not precisely measured by, the level of appropriable quasi rents. As these appropriable quasi rents increase, ex post bargaining opportunities between an external supplier and the focal firm escalate. Thus, like Williamson (1975, 1985), Klein et al. (1978) imply that the need to support the development of capability in an exchange drives the decision to integrate. Managers integrate because firm governance succeeds in promoting capability formation where market governance fails.

The explanation for firm boundary choice that derives from the knowledge-based view of the firm presents a message quite consistent with this original TCE logic. While TCE essentially defines the virtues of hierarchy as overcoming market failure, these scholars more fully articulate the governance virtues of hierarchy in generating capability (Kogut & Zander 1996; Conner 1991; Conner & Prahalad 1996; Teece, Pisano, & Shuen 1997). Thus, as Madhok (2002) argues, this “capability-based view of the firm, contends

---

6 Note the correspondence here is precise when referencing a dyadic exchange with dyadic investment. The logic becomes more complex when more than two actors or entities are involved.
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 markets simply cannot” (p. 536). Some of this literature focuses on the greater capacity to direct actors within the firm rather than convincing or educating them as to how to act (Conner & Prahalad 1996; Demsetz 1988). Other literature focuses more on the social advantages of hierarchy in promoting common identity, community, and social norms of exchange (Kogut & Zander 1996; Nahapiet & Ghoshal 1998). While these scholars introduce valuable new logic to explain the virtues of firm governance in generating capability, the arguments are very consistent with the broad conceptual apparatus of TCE. Indeed, the beginnings of these arguments are can be found in Williamson’s early discussion of internal adaptation and organizational atmosphere (e.g., 1975: pp. 25-26, 37-39). Thus, while TCE has placed heavy emphasis on articulating the failings of the market which hierarchy overcomes and is admittedly deficient in description of virtues of hierarchy, the capability scholars emphasize the virtues of the firm, but ignore or dismiss the causes of market failure that necessitate the firm, or the limits of firms that necessitate the market (Foss 1996a, 1996b). While TCE is clearly deficient in articulating the virtues of hierarchy, the capabilities based logic fails to explain the failures of firm organization that give rise to markets. While a complete theory of the firm clearly requires a more compelling articulation of all costs and benefits associated with alternative governance arrangements, it is clear that from the outset TCE has focused on how governance choices create and protect appropriable quasi rents, which is the outcome of capability development.
In summary, therefore, transaction cost logic can be understood to argue that efforts to generate unique capability through exchange with an outside supplier produce hazardous exchange conditions that promote integration. We argue that correctly interpreted, this logic implies that the desire to generate unique capability drives the decision to integrate. Efforts to generate unique capability through market exchange, on the other hand, generate significant appropriable quasi-rents that markets have difficulty allocating efficiently. Their presence generates high contracting costs. Integration eliminates these costs (though it introduces a new set of costs) thereby promoting capability development and protecting it from expropriation.

**Boundary Choices and the Origin of Capability**

If we think of transaction governance as the problem of governing capability development, the difficulty in using comparative capability logic alone to explain boundary decisions becomes apparent. It also highlights the difficulty faced in interpreting recent empirical literature that attempts to directly compare capabilities and transaction cost explanations of firm boundaries. A common conclusion in this literature is that a transaction is organized in a particular way (e.g., internalized or conducted through the market) at a given point in time more for comparative capabilities reasons than for transaction cost reasons. Interpreting observations of boundary choices in this static way, however, is problematic. The difficulty is that identifying comparative

---

7 As noted above, the knowledge based theory of the firm literature highlights the distinctive coordination advantages of hierarchy rather than hierarchy’s capacity to avoid market failure (Grant 1996; Conner & Prahalad 1996; Kogut & Zander 1996; Madhok 2002). But, of course, this may merely be describing two sides of the same coin. While TCE explains why markets can’t coordinate well due to market hazards and but is less developed regarding what it is that hierarchies can do, the KBV explains why hierarchies are superior in coordination, but it doesn’t fully explain why markets cannot replicate this (Foss 1996a, 1996b).
capability does not explain the origin of that capability, nor the role that transaction costs might have played in prior boundary decisions involved in its development. Moreover, as we discuss later, it fails to explain the persistence of a decision to integrate once the capability is formed.

The distribution of specialized capabilities across firms and their buyers and suppliers at a particular point in time reflects a series of past decisions by these firms to either develop or not to develop capabilities internally. Thus, the possession of a capability today reflects a choice to internally develop (or purchase) that capability yesterday. These decisions, we argue, were likely driven by comparative governance or transaction cost considerations. Consider the following example. A firm decides to internalize an activity at time 1 because performing this activity with the desired level of capability requires highly idiosyncratic investments—investments that suppliers are reluctant to make in the absence of carefully crafted safeguards. Due to the high costs of contractually creating and enforcing these safeguards, the firm chooses to integrate this capability development. As these specific investments are made over time, the firm develops the desired, superior capability to perform the activity, so that by time 2, the capability is fully developed, leaving no outside supplier with a comparable capability. Thereafter, the firm continues to be integrated.

Which theory, TCE or capabilities logic, best explains the firm’s boundary choice at time 2 in this example? A static analysis of the decision at time 2 would assign a small role to transaction costs and a large role to capabilities logic, yet this would be misleading if interpreted to imply that transaction costs were never important in the boundary choice made by the firm. In fact, the capability difference between the firm and its (potential)
suppliers only arose from an integration decision made at time 1, which was driven by transaction cost considerations. The integration decision at time 2 therefore cannot be interpreted as following simply from comparative capability, and not transaction cost logic. Yet some recent studies of make-or-buy decisions promote this kind of conclusion.

The question of why a capability was originally developed internally is different from the question of why it continues to be internally governed. At some initial point in time, before a specialized capability has been formed by a buyer, a simple comparative capabilities test may well suggest the need to exchange with an external provider with superior capability. A desire to generate unique capability through co-specialized investments, however, may cause the buyer to integrate the exchange, perhaps by purchasing the more capable external provider. Attempting to generate this unique capability through a market exchange generates high transaction costs. Once the unique capability is internally generated, and we apply the comparative capability test, that test clearly predicts the persistence of integration. However, the origin of the integration decision clearly does not hinge on comparative capability, but rather on the comparative efficiency of alternative governance forms in supporting capability development. Thus, while a static view might suggest that comparative capabilities alone can determine a firm’s boundary choices a given point in time, a dynamic perspective suggests that capabilities and transaction cost determinants are not independent of each other.

**Illustration: Ford Motor Company**

The dynamic interaction between transaction cost and capabilities explanations of firm boundaries we have just discussed is evident in boundary choices made by Ford
Motor Company during the early decades of the U.S. automobile industry. We suggest that some of Ford’s key early boundary choices were determined by transaction costs, that these choices then guided capability development, and that these capabilities in turn influenced later boundary choices.

Despite a well developed supplier base in the automotive industry, Ford chose to vertically integrate aggressively into automobile components as it introduced its moving assembly line in 1909. In order for this new manufacturing strategy to work, interchangeable parts had to be created, which in turn required highly firm-specific, single purpose machine tools — some of which were customized to Ford’s Model T. These tools included transmission testing devices, double-ended tools for pressing tubes in radiator fins, a tool for curling the heads of the gasoline tanks, radiator assembly and wheel painting machines, and the like. Most of these machine tools were designed by the 25-member Tool Design Department located in Ford’s Highland Park Plant (Nevins with Hill 1954: 456, 463-4).

The firm-specific investments required to create the machine tools necessary for assembly line production implies that high transaction costs played an important role in Ford’s decision to vertically integrate the production of components. Interestingly, Langlois & Robertson (1995), who examine this piece of Ford’s history in detail, do not entertain the possibility that transaction costs, as conventionally understood, played a role in Ford’s vertical integration decisions. Instead, they argue that it was the lack of capabilities among existing suppliers to produce the specialized machine tools that accounts for these decisions to vertically integrate. They argue that, “only the men of Ford understood the uses to which such machines would be put” (p. 53). This, however,
is precisely the logic that Williamson and Klein et al. articulated for why transaction costs drive vertical integration choices. Firms integrate transactions when they seek the development of a firm-specific capability that necessitates a transaction-specific investment(s) by supplier(s). By integrating the supplier(s), a firm avoids the need to contractually manage these specific investments required for capability development. In this case, the specific investments required for an external supplier to understand Ford’s products and technology, and then to design and produce the idiosyncratic machine tools Ford demanded, were likely too high to permit the efficient use of the market.

As time went on, Ford persisted in its vertical integration strategy. Indeed, through the 1920’s and early 1930’s, Ford remained highly vertically integrated even as most other companies vertically disintegrated (even while they increasingly adopted the moving assembly line). Katz (1977), for example, provides evidence that several companies, but not Ford, began vertically disintegrating after 1926. Argyres and Bigelow’s (2007) analysis of a sample of U.S. auto companies during the late 1920’s and early 1930’s also reveals a decline in the vertical integration level of the average firm. Why did Ford buck the trend? One possible answer is that Ford’s capabilities in producing the specialized machine tools necessary for assembly line production had become superior to existing suppliers’. By the mid-1920’s, Ford had at least a 15 year lead on suppliers in producing this kind of equipment, enough time to have moved down the learning curve—well ahead of available suppliers. Precision metal cutting may have been one of the key capabilities in which Ford was superior, and likely contributed to the large reductions in component costs Ford achieved after integration (Williams et al. 1993; Langlois & Robertson 1995: p. 52). It seems plausible, then, that Ford’s decision to
persist in its vertical integration strategy during the 1920’s and early 1930’s, even while competitors were actively disintegrating under similar assembly line production processes, resulted in part from past decisions to vertical integrate that were themselves driven to a significant degree by conventional transaction costs. Langlois and Robertson’s (1995) account of Ford’s development of metal stamping techniques follows a similar pattern: An initial decision to internalize production, in this case by acquiring a supplier named Keim, led to the development of new stamping capabilities that reinforced this integration decision through time (Langlois & Robertson 1995: pp. 53-54). Thus, in the case of Ford, transaction costs and capability development interacted over time in way that makes it difficult to attribute any given boundary decision to capabilities considerations that were independent of transaction costs considerations.

The Persistence of Boundary Choices

To this point we have argued that once we consider a capability’s origin and development, it becomes clear that comparative capabilities and transaction costs do not operate as independent explanations of firm boundary choices, but rather interact to affect those choices over time. A second, related question is whether capabilities considerations alone can explain the persistence of a decision to integrate or outsource. Thus, suppose that an outside supplier possesses capability superior to the firm’s internal capability, and the transaction in question is outsourced to that supplier. Under what conditions will this exchange continue to occur through the market with this external supplier? Or, suppose that an internal supplier possesses capability superior to all external suppliers. Under what conditions will that exchange remain within the boundaries of the firm? We contend
that comparative capabilities may determine the desired provider, but does not alone
determine the desired form of governing the exchange with that provider through time.
Transaction costs inevitably play a role as well.

Note first that the fact that firms routinely sell or spin off internal capabilities and
actively acquire or invest internally when they lack capability suggests that capabilities
alone cannot drive boundary decisions, at least over the long run. If firms desiring
capability can eventually acquire that capability, then lack of capability is not a sufficient
for the persistence of a decision to outsource. Conversely, if a firm can choose to keep or
sell off a superior capability, a decision to continue to internalize production cannot be
explained solely by the possession of the capability today.

Acquiring or selling a capability, however, can sometimes be difficult to
accomplish in the short run. Indeed, scholars advocating the comparative capabilities
logic have emphasized these difficulties in explaining why capabilities considerations
alone loom large in boundary decisions (Barney 1999; Langlois 1992; Zott & Amit
2006). Thus, Zott and Amit (2006) argue that time compression diseconomies (Dierckx
and Cool 1989) restrict the capacity of firms to internally develop capabilities that are
accessible on the market. Langlois (1992) similarly argues that firms often cannot acquire
a capability on the same time scale as make-or-buy decisions for particular transactions.
Thus, consider a firm that is developing a new product, but lacks the capability to
produce a highly unique and highly valuable component or tool inhouse. While this
capability may be easily assembled from relatively generic assets and easily obtained in
the marketplace, the time to develop or acquire this capability internally may be
considerable. In the short run, the firm may rely on outside suppliers, even if the unique nature of the component puts the firm at risk of hold-up.

One issue with these arguments is that transaction costs of various kinds are likely to be at work in helping to cause development lags and time compression diseconomies, implying that capabilities considerations are not operating alone. For example, such lags may well reflect costs of searching for the appropriate providers of labor, capital, and skills needed for the capability development, as well as negotiation and enforcing contractual agreements with them. These are the three categories of transaction costs emphasized by Coase (1937) and North (1990).

A second issue is whether anything precludes the firm from simply acquiring the capability outright by purchasing the firm (or a portion thereof) which possesses it. Barney (1999) suggests a variety of impediments to simply purchasing capability. For example, the process of integration itself may diminish the quality of the acquired capability. Acquiring capabilities may also be difficult to reverse and thus limit flexibility. Barney (1999) also mentions legal impediments to acquiring capability. In addition, the desired capability may be inextricably bound with other unwanted capabilities, or similarly, may have capacity to contribute beyond the input needs of the acquirer and thus be underutilized post acquisition (e.g., Penrose 1959).

Each of these impediments, however, is likely rooted in transaction cost problems. Problems of integrating acquisitions or effectively leveraging unrelated capabilities that accompany the focal capability can stem from the organizational costs involved in attempting to standardize incentive and governance arrangements across existing and acquired units (e.g., Williamson 1985: p. 158). Problems of irreversibility and of selling
unwanted capabilities can be caused by high transaction costs involved in selling business units in thin markets in which information about the value of the unit is asymmetric as between buyer and seller (Williamson’s (1975) “small numbers bargaining” and “information impactedness” problems). Legal or contractual impediments to acquisition can result from poorly specified property rights, leading to the kinds of transaction costs also emphasized by Coase (1960), Barzel (1982) and North (1990). Thus, while a firm’s lack of a key capability may be a proximate cause of a decision to outsource at a given point in time, the reasons why the firm lacks that capability in the first place are likely to include transaction costs in some form.

Just as the lack of a capability does not (without transaction costs) explain a decision to persist in outsourcing, the possession of a comparative capability does not by itself explain the decision to retain that capability and continue to internalize production. Note first that retaining a superior capability inhouse indefinitely is not necessarily efficient. For example, because buyers are generally reluctant to purchase inputs from competitors for transaction cost reasons (Williamson 1975: pp. 16-19; Chen 2005), retaining a superior capability inhouse, by insulating it from the demand of multiple buyers, can cause it to atrophy. AT&T’s decision to sell off its equipment division (Kirkpatrick 1995), and General Motors’s decision to sell off its Delphi parts division (Tait 1999), and are examples that appear to reflect these kinds of concerns. As we noted, firms routinely sell off capabilities, and therefore the mere possession of a capability cannot explain the persistence of an integration decision. On the other hand, in cases where selling capabilities is difficult to accomplish in the short run, once again transaction costs are likely to be lurking. The transaction costs involved in selling
capabilities are of course the same as those involved in purchasing them on the market, and involve small numbers bargaining problems, information impactedness, asset specificity, poorly defined property rights, and the like.

In summary, we argue that just as capabilities considerations alone do not completely explain boundary decisions at a point in time, they do not alone explain the persistence of those decisions. Instead, capabilities considerations go hand-in-hand with transaction costs and are ultimately rooted in them. Seen from this perspective, our argument is not that capabilities do not matter in integration decisions, but that they matter greatly, precisely because their development and exchange so often involves high transaction costs. In the remainder of the paper we articulate an approach to firm boundaries that integrates the logic of both the capabilities literature and transaction cost economics, and takes account of the dynamic interactions of between these two sets of considerations.

**Toward An Integrative Approach to Firm Boundaries**

There is growing consensus in the strategy literature that firms acquire capability and the resulting rents or positions of advantage by assembling sets of unique and complementary resources, activities, or assets. Thus, Rumelt (1984) argues that “a firm’s competitive position is defined by a bundle of unique resources and relationships” (pp. 557-558). Other scholars highlight complementarity or superadditivity among a firm’s resources, activities, or assets, where complementarity is defined as the “property that doing more of any subgroup of activities raises the marginal return to the other activities” (Milgrom & Roberts 1990). Consistent with this logic, Montgomery and Wernerfelt
(1988) note that through “superadditive productivity” among actors, firms generate rents which specific actors are unable to appropriate themselves. Amit and Shoemaker (1993: 39) note that rent-generating capabilities arise from resource complementarity. Dierickx & Cool (1989) point to the role that the “interconnectedness of asset stocks” plays in generating privileged asset positions which generate rents. Ghemawat (2005) argues that competitive advantage stems from “an integrated set of choices about activities,” noting that “a firm whose choices do not fit together well is unlikely to succeed” (p. 131). Lippman & Rumelt (2003) argue that the field of strategy broadly “concerns the creation, evaluation, manipulation, administration, and deployment of unpriced specialized scarce resource combinations.” (p.1069). 8

Thus, while there are differences in language as to the elements that are complementary, superadditive or interconnected (i.e. activities, assets, or resources), the broad concept is quite consistent. Capability, competitive advantage, or rent generating positions derive from the presence of complementarity. Managers who can uniquely identify these complementary bundles and then effectively assemble them without revealing the value of the bundle to others during the process of assembly are able to build rent-generating positions or capability. Our interest, however, is in explaining how firms determine the governance form through which these assets, activities, or resources are accessed. In particular, which of the activities, assets, or resources which comprise or

---

8 Barney (1986) and others point out that these complementary resource or activity bundles must be uniquely perceived at the time the component resources are acquired, or else competition in factor markets will correctly price these assets and thereby consume any rents.
help form a complementary bundle should be owned by the firm and which can be effectively outsourced?  

We contend that the question of optimal governance for a given activity requires a clear understanding of the relationships among complementarity, co-specialization, and “elemental” capability—i.e., capability that is an element of a bundle of complements rather than the capability that results from the bundle as a whole.  

Consistent with prior work, our contention is that while the presence of complementarity or fit among activities is necessary to rent generation, it is not sufficient to explain the governance choice (Teece 1986). Instead, it is the presence or absence of co-specialization between a focal activity and other elements in that bundle which encourages integration, where co-specialization is defined, consistent with Milgrom and Roberts (1990), as complementarity among elements that is unique in some respect. The distinction here is that one can easily have an activity or asset that is a critical part of a superadditive bundle of elements that generates rents, but is in no way co-specialized to these other activities. In this case, the focal asset or activity may be quite generic, with many different firms capable of providing these same assets or activities in precisely the same way. There is no transaction cost problem associated with using the market to manage this complementarity. Any effort on the part of the outside firm to appropriate the rents generated by this complementary bundle of activity choices, of which their asset is a part, would result in their replacement by another firm possessing identical assets.

---

9 While the central question in TCE has long focused on how boundary choices influence a firm’s capacity to generate, protect, and appropriate rents, the analysis has been dyadic, focusing on the governance of an exchange between a focal firm and a distinct activity or asset. Thus, one of the challenges in integrating TCE and capabilities logic is attempting to explain boundary choices that are inherently based on multi-lateral exchanges, involving bundles of complementary activities, assets, and resources.

10 Presumably, there is a bundle of complementary activities, assets, or resources that underlies this elemental capability as well.
Suppose instead that a focal activity in a complementary bundle is also co-specialized to this bundle of complements, i.e. it is complementary in some unique way. The degree of unique complementarity or co-specialization can be defined as the value of the bundle of complements with the focal provider providing the focal activity relative to the maximum value of this bundle achievable with some other provider providing the focal activity instead. This difference defines the appropriable quasi rents generated by the bundle when the focal provider participates.

This definition follows from recent game-theoretic approaches to conceptualizing the role that resources play in generating rents (Brandenburger & Stuart 1996). Lippman and Rumelt (2003), for example, treat resources as “players” in a game. Co-specialized resources are defined to be coalitions of multiple players that yield a higher payoff to each coalition member than the member could earn in alternate coalitions. This conceptualization of co-specialized resources is, however, agnostic about the means – the governance form -- by which a coalition of resources is actually achieved. We contend that if producing rents from a complementary bundle of resources or activities involves a significant degree of co-specialization, then safeguards are required to protect the exchange and thereby prompt such investments. Thus, when exchanges require substantial co-specialization to generate this complementarity, integration is more likely. In this manner, complementarity *per se* is a necessary but not sufficient condition to explain integration.

To illustrate the relationships among capability, co-specialization, and complementarity, consider a bundle of activities that are highly complementary in the sense that carrying out one activity greatly increases the marginal return from performing
the other activities. Assume that a focal firm integrates all of the activities but Activity A. What role does comparative capability play in determining whether the firm integrates Activity A as well?

We contend that the focal firm will choose to both exchange with and be prompted to own the most capable provider of Activity A if this superior capability generates unique complementarity or co-specialization with the bundle of complements. In this case, the most capable provider of Activity A is more valuable to this bundle of activities than it is to any other bundle. A decision to exchange with any other less capable provider dramatically lowers the value which this bundle of complements can generate. Therefore, failure to integrate the highly capable Activity A provider exposes the firm to potential hold-up of appropriable quasi rents by this provider. Note that capability assessment here is always context specific in that the magnitude of comparative capability for conducting an activity is defined by the unique value it contributes to a particular bundle of activities. Thus, the most capable provider of Activity A for a given bundle of complements is the one which generates the most value with that bundle. In this regard, co-specialization and comparative capability are synonymous when analyzed in the context of a particular bundle of complements; the greater the unique value generated by an activity, the more co-specialized it is. Moreover, the scope of appropriable quasi rents generated by a capability is measured as the value of the bundle of complements created with the most capable provider of Activity A relative to the value of this bundle with the next most capable provider. Thus, when the unique value added by a capability provider to a particular bundle of complements is large, integration is likely.
We can think of the RBV, then, as dealing with the question of which resources complement each other to produce a competitive advantage, while TCE deals with the question of which of these complementary resources will come under common ownership of the firm, and which will be independently owned. Our contribution here is to point out that it is not comparative capability or complementarity that alone determines the boundary choices of a focal firm. Comparative capability and complementarity only determine the desired providers. The preferred form of governance with respect to that provider is determined by the degree to which the capability in question is co-specialized to the bundle of resources or activities of the firm in question.

**Illustration: Disney**

Disney’s history in sourcing its animation activity over several decades illustrates the relationships among complementarity, co-specialization, and firm boundaries that we The role that complementarity played in generating competitive advantage for Disney was evident to its management quite early in the company’s history. For instance, Figure 1 provides an activity or synergy map connecting Disney’s activities that was drawn by company managers in the late 1950s. Disney management clearly perceived strong complementary relationships among its investments in films, consumer products, theme parks, books, magazines, and music. Moreover, Disney identified film production, particularly animated film production, as the activity most complementary with these other activities (Collis & Montgomery 1995; Rukstad, Collis & Levine 2001).

Considering the relationships among Disney’s key capabilities and resources is useful for illustrating our argument concerning resource co-specialization and boundary
choices. First, one could make the case that no other bundle of complementary activities gains more value from superior capability in animated film production than the bundle of complements owned or assembled by Disney. In this sense, Disney’s other activities have not only been complementary with, but uniquely complementary with (or co-specialized to) a capability in animation. Our argument would then predict that attempting to access this type of highly co-specialized animation capability through contracts rather than internal organization would leave Disney vulnerable to the appropriation of quasi-rents.

By the late 1990’s, Disney no longer possessed the superlative capability in animation—that designation belonged to Pixar, which had developed the most advanced computer-based animation capability in the film industry. Disney initially accessed this capability through a contractual relationship with Pixar that began in 1991 with a single-film deal. Following the success of that first film, *Toy Story*, a multi-film deal was signed. Within a few years, Disney’s strong dependency on Pixar became clear, and Pixar found itself in a position to appropriate a substantial portion of the quasi-rents generated in this bundle of complements.\(^{11}\) Pixar attempted this appropriation during contentious contract renegotiations lasting ten months, and leading Pixar to terminate the relationship in 2004, (*Wired*, 1/29/04). Consistent with our prediction, Disney soon thereafter entered negotiations to acquire Pixar and thereby its uniquely complementary capability. The acquisition was completed in January 2006 for $7.4 billion in stock, considered by analysts to be a very high price. The key point to be drawn from this example is that it

\(^{11}\) Pixar was apparently less dependent on Disney than *vice versa*, because other film studios possessed some of the assets (such as access to worldwide theater distribution) that Pixar lacked but required to appropriate returns from its investments in computer animation. Indeed, Pixar was also negotiating with Warner Bros. and Fox while it was renegotiating with Disney.
was not simply superior capability that drove the integration decision in this case. Instead, it was capability that was uniquely complementary with or co-specialized to the other activities of the acquiring firm.

**Conclusion**

In our view, the growing strategy literature on the determinants of firm boundaries contains a misleading thread. A false dichotomy between transaction cost and capabilities theories of firm boundaries has emerged. As contributors to this early literature, we acknowledge our own contributions to the misunderstanding. In this paper, we have argued that TCE, properly understood, is vitally concerned with the question of which governance choices facilitate the development of which kinds of capabilities. Moreover, we argue that if capability differences between a firm and its potential suppliers play a key role in determining the firm’s governance choices, it is more than likely that transaction costs lie somewhere at the roots of these differences. Therefore, scholars should cease to assume the existence of a meaningful distinction between transaction cost and capabilities-based theories of firm boundaries – one that can be used to drive empirical research on the determinants of those boundaries. Instead, scholars should treat capabilities considerations as inextricably intertwined with transaction cost logic, and should seek to analyze aspects of this complex interaction.

One way to research this interaction is to investigate how governance choices impact the rate and efficiency of capability formation. The knowledge-based theory of the firm literature takes up this agenda directly, highlighting the virtues of the firm in supporting knowledge flows and knowledge creation (e.g., Grant 1996; Conner &
Prahalad 1996). The agenda that remains here is to better understand the limits to firms’ knowledge creation abilities, as well as to understand the roles of various organizational features in moving organizations toward these limits. For example, with respect to internal organization, what kinds of authority and incentive structures facilitate the development of new knowledge within the firm (e.g., Nickerson & Zenger 2004)? How does informal organization stimulate the emergence of new knowledge within the firm by reducing various kinds of organizational costs (e.g., Argyres & Mui 2007)? With respect to alliances, what kinds of contractual safeguards facilitate the transfer of tacit knowledge and encourage the development of desired routines (e.g., Oxley 2003)? Research aimed at these kinds of questions will no doubt turn up insights into the relationships between governance choice and capabilities development.

Another set of issues related to the interaction of transaction costs and capabilities revolves around the question of how firms learn to govern their internal and external relationships over time. Whereas capabilities-based theories of the firm have tended to emphasize that the key firm capabilities are technological in nature, recently scholars have begun to explore the possibility that firms can develop capabilities for governing activities in ways similar to those in which they develop production capabilities (e.g., Child 1999; Anand & Khanna 2000; Azoulay & Shane 2001; Dyer & Singh 1998; Mayer & Argyres 2004). Research on capability for governing aims to uncover insights about how firm boundaries evolve as firms, particularly those in emerging industries, hone their capabilities for governing internal and external exchange relationships.
Thus, further research on the way firms govern their learning activities, as well as how they learn to govern their activities, promises to deepen our understanding of the interactions between transaction cost and capabilities determinants of firm boundaries.
References


Argyres, N. and Bigelow, L. 2007. Competitive Positioning, Dominant Design and Vertical Integration Over the Industry Lifecycle. manuscript, Boston University School of Management.


Kirkpatrick, D. 1995. AT&T Has the Plan. Fortune, October 16, p. 84.


Figure 1